```
                                    GenCore version 4.5
                               Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:            May 25, 2001, 03:04:44 ; Search time 125.73 Seconds
                                                      (without alignments)
Title:             US-09-403-269-12_Copy_1_1404          6518.961 Million cell updates/sec
Perfect score:     1404
Sequence:          1 tccaagctgaattcctatag..........gcagccggcaaagcacaac 1404

Scoring table:     IDENTITY.NUC
                   Gapop 10.0 , Gapext 1.0

Searched:          678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters:       1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:   Minimum Match 0%
                   Maximum Match 100%
                   Listing first 45 summaries

Database :         1: N_Geneseq_0401:*
                   2: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
                   3: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
                   4: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
                   5: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
                   6: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
                   7: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
                   8: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
                   9: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
                   10: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
                   11: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
                   12: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
                   13: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
                   14: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
                   15: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
                   16: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
                   17: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
                   18: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
                   19: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
                   20: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
                   21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
                   22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*
```

SUMMARIES

| Result NO. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 1404 | 100.0 | 3085 | 19 | V62688 | Bovine glucuronyl |
| c 2 | 35.6 | 2.5 | 237326 | 19 | V57903 | Hereditary haemoch |
| c 3 | 35 | 2.5 | 1366 | 20 | 227743 | Human DNA marker c |
| c 4 | 35 | 2.5 | 5059 | 20 | X84332 | Stealth virus nucl |
| c 5 | 34.4 | 2.5 | 2248 | 20 | C69113 | Human secreted pro |
| c 6 | 33.8 | 2.4 | 571 | 20 | X20393 | Borrelia burgdorfe |
| c 7 | 33.8 | 2.4 | 9542 | 20 | X20260 | Borrelia burgdorfe |
| c 8 | 33.6 | 2.4 | 283 | 21 | C94048 | Cat flea hindgut a |
| c 9 | 33.6 | 2.4 | 1231 | 21 | C39833 | Arabidopsis thalia |
| c 10 | 33.6 | 2.4 | 1392 | 19 | V44080 | mu-NEO-r DNA fragm |
| c 11 | 33.6 | 2.4 | 1534 | 16 | Q92554 | Bovine conglutinin |

```
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

| | | | | | | |
|---|---|---|---|---|---|---|
| c 12 | 33 | 2.3 | 1644 | 20 | X60775 | Soybean cytochrome |
| 13 | 32.6 | 2.3 | 3970 | 21 | A97911 | L. mesenteroides a |
| 14 | 32.6 | 2.3 | 6204 | 21 | A97917 | L. mesenteroides a |
| 15 | 32.6 | 2.3 | 7387 | 21 | A97914 | L. mesenteroides a |
| 16 | 32.6 | 2.3 | 9321 | 21 | A97904 | L. mesenteroides a |
| 17 | 32.6 | 2.3 | 49999 | 20 | Z23903 | L. mesenteroides a |
| c 18 | 32.4 | 2.3 | 72604 | 20 | Z10752 | Human LOBO homolog |
| c 19 | 32 | 2.3 | 1334 | 21 | A39409 | Genomic sequence o |
| 20 | 32 | 2.3 | 1958 | 19 | V52493 | Rice SYR2 homologu |
| c 21 | 32 | 2.3 | 2423 | 21 | C98795 | Streptococcus pneu |
| 22 | 31.8 | 2.3 | 1661 | 17 | T45120 | Human pancreatic c |
| c 23 | 31.8 | 2.3 | 1661 | 19 | V03875 | HTm4 protein codin |
| 24 | 31.8 | 2.3 | 13339 | 19 | V44234 | HTm4 gene. Homo s |
| 25 | 31.6 | 2.3 | 624 | 21 | F07873 | Lettuce resistance |
| c 26 | 31.6 | 2.3 | 2547 | 21 | A74993 | Fusarium venenatum |
| 27 | 31.6 | 2.3 | 2773 | 21 | Z65009 | cDNA encoding a hu |
| 28 | 31.6 | 2.3 | 4473 | 22 | F44671 | Membrane-bound pro |
| c 29 | 31.4 | 2.2 | 1291 | 21 | C30019 | Novel protein kina |
| 30 | 31.4 | 2.2 | 1238 | 21 | 252513 | Human secreted pro |
| c 31 | 31.4 | 2.2 | 1365 | 21 | C39019 | Arabidopsis thalia |
| 32 | 31.4 | 2.2 | 1330 | 21 | C49301 | Human CD2 cytoplas |
| 33 | 31.4 | 2.2 | 1333 | 21 | C36038 | Arabidopsis thalia |
| c 34 | 31.4 | 2.2 | 1365 | 21 | A49460 | Arabidopsis thalia |
| 35 | 31.4 | 2.2 | 1447 | 20 | X85015 | Human secreted pro |
| c 36 | 31.4 | 2.2 | 1486 | 21 | X135590 | Enterococcus faeca |
| 37 | 31.4 | 2.2 | 1627 | 16 | F18228 | Lung cancer associ |
| 38 | 31.4 | 2.2 | 1627 | 16 | Q95058 | Chicken TEF-1A CDN |
| 39 | 31.4 | 2.2 | 1627 | 19 | V37721 | Chicken TEF-1A CDN |
| 40 | 31.4 | 2.2 | 1666 | 16 | Q95059 | Recombinant transc |
| 41 | 31.4 | 2.2 | 1666 | 19 | V37722 | Chicken TEF-1B CDN |
| 42 | 31.4 | 2.2 | 3463 | 21 | C77382 | Recombinant transc |
| 43 | 31.4 | 2.2 | 51259 | 21 | X83007 | Partial mouse WRN |
| 44 | 31.2 | 2.2 | 1664976 | 21 | V21209 | Methanococcus jann |
| 45 | 31.2 | 2.2 | 1551 | 21 | C77099 | Human ORFX ORF2654 |
| | 31.2 | | 1646 | 9 | N82246 | Rice storage prote |

ALIGNMENTS

```
RESULT  1
ID   V62688   standard; cDNA; 3085 BP.
XX
AC   V62688;
XX
DT   15-FEB-1999  (first entry)
XX
DE   Bovine glucuronyl C5-epimerase cDNA.
XX
KW   Glucuronyl C5-epimerase; cattle; D-glucuronic acid;
KW   L-iduronic acid; heparin; heparan sulphate; ss.
XX
OS   Bos taurus.
XX
FH   Key             Location/Qualifiers
FT   CDS             73..1407
FT                   /*tag= a
XX
PN   WO9848006-A1.
XX
PD   29-OCT-1998.
XX
PF   17-APR-1998;  98WO-SE00703.
XX
PR   18-APR-1997;  97SE-0001454.
XX
PA   (LIJJ/) LI J.
PA   (LIND/) LINDAHL U.
XX
PI   Li J,  Lindahl U;
XX
DR   WPI; 1998-583655/49.
DR   P-PSDB; W79263.
```

XX

PT  DNA sequence coding for mammalian glucuronyl C5-epimerase and
PT  functional derivatives - capable of converting D-glucuronic acid to
PT  L-iduronic acid in the synthesis of heparin and heparan sulphate

XX

PS  Claim 2; Page 18-19; 26pp; English.

XX

CC  This cDNA sequence includes a claimed coding region that encodes
CC  bovine glucuronyl C5-epimerase (see W79263), an enzyme that
CC  catalyses the conversion of D-glucuronic acid (GlcA) to L-iduronic
CC  acid (IdoA).  To isolate the sequence, highly purified epimerase
CC  from bovine liver was subjected to digestion with a lysine-specific
CC  protease.  A DNA probe was generated from bovine liver cDNA by PCR
CC  using primers (see V62689-91) based on an isolated peptide (see
CC  W79270).  This was used to screen a bovine lung lambda gt10 library,
CC  and a hybridising clone was identified and sequenced. An additional
CC  12 bp of 5' sequence was obtained from a clone isolated from the
CC  library by PCR using an epimerase-specific primer (see V62692).
CC  The invention relates to isolated or recombinant DNA sequences for
CC  a mammalian (including human) glucuronyl C5-epimerase or its
CC  functional derivative.  Recombinant expression vectors and
CC  transformed host cells are also claimed.  The nucleic acid and
CC  vector can be used for the recombinant production of the enzyme.
CC  Glucuronyl C5-epimerase is useful for converting GlcA to IdoA in
CC  the biosynthesis of heparin and heparan sulphate.

XX

SQ  Sequence 3085 BP; 903 A; 625 C; 654 G; 903 T; 0 other;

Query Match             100.0%;  Score 1404;  DB 19;  Length 3085;
Best Local Similarity   100.0%;
Matches 1404;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

```
Qy          1 tccaagctgaattcttcattagctattccaaagtctatgcacagagagccccttatcacct 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 tccaagctgaattcttcatagctattccaaagtctatgcacagagagccccttatcacct 60

Qy         61 gatggtgtgttcttatgtccttgaaggctacacaatgtggaagtccgagacagagtcaagtgc 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 gatggtgtgttcttatgtccttgaaggctacacaatgtggaagtccgagacagagtcaagtgc 120

Qy        121 ataagtgggttgaaggtgtaccttttatctcacacagtggggacctcaaggctattctac 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ataagtgggttgaaggtgtaccttttatctcacacagtggggacctcaaggctattctac 180

Qy        181 ccaatccagattgcacagtatggttaagtcactcacgacaagaatctaactgaaaacc 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ccaatccagatgtcacagtatggttaagtcactcacgacaagaatctaactgaaaacc 240

Qy        241 cccatatagaggtatatgaaacacagaagaacagcaagcaagcccaatgac 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 cccatatagaggtatatgaaacacagaagaacagcaagcaagcccaatgac 300

Qy        301 tggactgtgcccaagggctgcttattcggctagtgtgtggcctgataagttcaccaat 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 tggactgtgcccaagggctgcttattcggctagtgtgtggctgataagttcaccaat 360

Qy        361 gttaaacagttcattgcccaagaaacagtgaaggtgtatcgttgggaacaca 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 gttaaacagttcattgctccagaaacagtgaaggtgtatcctgcaactgggaacaca 420

Qy        421 aaagatttattattcattgacctcaagtctcttaccaaatggaagcgtgtctgtggtt 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 aaagatttattattcattgacctcaagtctcttaccaaatggaagcgtgtctgtggtt 480

Qy        481 ctggagacgacgaaaagaatcagctctcactgtatcaaataacccagcta 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ctggagacgacgaaaagaatcagctctcactgtatcaaataacccagcta 540

Qy        541 attgctcttaaagaaagaagaacatactatggcatccggccagaacatcatggagagcaca 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 attgctcttaaagaaagaagaacatactatggcatccggccagaacatcatggagagcaca 600

Qy        601 gttaacccggacctggtcaactgacctcaggagaacctgaagaggagtgggtctttccaacacaaagct 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 gttaacccggacctggtcaactgacctcaggagaacctgaagaggagtgggtctttccaacacaaagct 660

Qy        661 gtcaagccaacaagaaataatgcccaagaaaggtgttaggtggttgaattgcagaaagggaaggggc 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 gtcaagccaacaagaaataatgcccaagaaaggtgttaggtggttgaattgcagaaagggaaggggc 720

Qy        721 ttccttgacaacattaccatcctacacagcaccacagcccacatggctgccttcttcgttcggtcaagt 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ttccttgacaacattaccatcctacacagcaccacagcccacatggctgccttcttcgttcggtcaagt 780

Qy        781 gactggctggtgaggaaccaggagagagaaagcggctggccgattatggtgacccgtaag 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 gactggctggtgaggaaccaggagagagaaagcggctggccgattatggtgacccgtaag 840

Qy        841 ttaggggaaggcttcaagtcttagagacgcaaggggtactggccaagtccaagggccaa 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ttaggggaaggcttcaagtcttagagacgcaaggggtactggccaagtccaagggccaa 900

Qy        901 gccatttctacattagtcaggggctatctcttaacaaaagcacatatattcctcaattca 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 gccatttctacattagtcaggggctatctcttaacaaaagcacatatattcctcaattca 960

Qy        961 gcttaagggcaaacgcccccttacaagttcctgcaagagcagcatgaagtcaaggctgtg 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 gcttaagggcaaacgcccccttacaagttcctgcaagagcagcatgaagtcaaggctgtg 1020

Qy       1021 tttatgaataaacatgactggttgtaagaagaatatccaactacacactagctcttttgtttta 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 tttatgaataaacatgactggttgtaagaagaatatccaactacacactagctcttttgtttta 1080

Qy       1081 aatggcttttatgtattcttttaattgggctgtgtatgacttaaaagaaactgcaggggaaaa 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 aatggcttttatgtattcttttaattgggctgtgtatgacttaaaagaaactgcaggggaaaa 1140

Qy       1141 ctcgggaagagaggtgtcctttgtatgagcgtggcatggaatgaagccaatgctc 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ctcgggaagagaggtgtcctttgtatgagcgtggcatggaatgaagccaatgctc 1200

Qy       1201 cccttgtacgacactggctcaggaaccatctatgacctccggcactctgctggcatt 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 cccttgtacgacactggctcaggaacctatgacctccggcactctgctggcatt 1260

Qy       1261 gccccccaacctgctgggactatcacaccaatcaatcaaactgcagctgctt 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 gccccccaacctgctgggactatcacaccaatcaatcaaactgcagctgctt 1320

Qy       1321 agcaccattgatgagtcccaatcttcaaagaaattgtcaagaggtggaagagctacctt 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 agcaccattgatgagtcccaatcttcaaagaaattgtcaagaggtggaagagctacctt 1380

Qy       1381 aaaggcagccgggcaaagcacaac 1404
              ||||||||||||||||||||||||
Db       1381 aaaggcagccgggcaaagcacaac 1404
```

RESULT  2

ID   V57903/c
AC   V57903 standard; DNA: 237326 BP.
XX
AC   V57903;
XX
DT   21-DEC-1998 (first entry)
XX
DE   Hereditary haemochromatosis subregion from an HH affected individual.
XX
KW   Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
KW   diagnosis; iron metabolism; NPT3; NPT4; RoRet; BTF1; BTF2; BTF3;
KW   BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW   type 1 sodium transport gene; ss.
XX
OS   Homo sapiens.

```
XX
PN   WO9814466-A1
XX
PD   09-APR-1998
XX
PF   30-SEP-1997;  97WO-US17658.
XX
PR   07-MAY-1997;  97US-0852495.
PR   01-OCT-1996;  96US-0724394.
XX
PA   (PROG-) PROGENTIOR INC.
XX
PI   Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ;
PI   Tsuchihashi Z, Wolff RK;
XX
DR   WPI: 1998-240014/21
XX
PT   Hereditary haemochromatosis gene products - used to develop products
PT   for the diagnosis and treatment of hereditary disorders in iron
PT   metabolism
XX
PS   Claim 1; Fig 9; 209pp; English.
XX
CC   The present invention describes hereditary haemochromatosis gene
CC   products from the human haemochromatosis gene. The present sequence
CC   represents a hereditary haemochromatosis subregion from an hereditary
CC   haemochromatosis (HH) affected individual. Also described is a
CC   method to determine the presence or absence of the common hereditary
CC   haemochromatosis (HFE) gene mutation in an individual comprising:
CC   (a) providing DNA or RNA from the individual; and (b) assessing the
CC   DNA or RNA for the presence or absence of a haplotype or genotype where
CC   the presence or absence of the haplotype genotype indicates the likely
CC   presence of the HFE gene mutation in the genome of the individual. The
CC   HFE gene sequences from the present invention can be used to develop
CC   products for use in the diagnosis and treatment of HFE. The present
CC   invention also describes BTF genes, which are homologues of the milk
CC   protein butyrophilin (BT), and can be used in the production of agonists
CC   and antagonists of BT function. Also described are: (1) a RoRet gene
CC   which can be used to develop products for the study, diagnosis and
CC   treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC   which are homologues of a type 1 sodium transport gene, and can
CC   similarly be used for hypophosphatemia.
XX
SQ   Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T; 0 other;

Query Match             2.5%;  Score 35.6;  DB 19;  Length 237326;
Best Local Similarity   60.2%;
Matches   59;  Conservative    0;  Mismatches  21;  Indels    0;  Gaps    0;

Qy         9 gaattctcatagctattccaaagtctatgcacagagaagcccctatcaccctgatggtgt 68
             || |||||| ||||| ||||| ||||| |||||||| ||| ||||| |||||||| |||||
Db     20379 GAGCTCACATAGCCHTTCCCAGTATATGTGCAGAGCTCCCTCTCHTCCCTCTTTTTTTTTT 20320

Qy        69 gtttatgtccttgaagctacacatgtggaagtccgag 106
             |||||| |||||||| ||| || |||||| |||||
Db     20319 TTTTCHTTTAAATAAATTATTTAANTTGGAGGACCAAG 20282


RESULT  3
Z27743

ID   Z27743 standard; DNA; 1366 BP.
XX
AC   Z27743;
XX
DT   23-DEC-1999 (first entry)
XX
DE   Human DNA marker clone S110.
XX
KW   Tandem repeat sequence; DNA isolation; intermediate tandem repeat;
KW   ITR sequence; pentanucleotide tandem repeat; stutter artifact;
KW   DNA typing; DNA profiling; linkage analysis; criminal justice;
KW   paternity testing; animal lineage analysis; microsatellite loci;
KW   polymorphism detection; ds.
XX
OS   Homo sapiens.
XX
PN   WO9940194-A1
XX
PD   12-AUG-1999
XX
PF   04-FEB-1999;  99WO-US02345.
XX
PR   04-FEB-1998;  98US-0018584.
XX
PA   (PROM-) PROMEGA CORP.
XX
PI   Schumm JW, Bacher JW;
XX
DR   WPI: 1999-590696/50.
XX
PT   Isolating DNA containing intermediate tandem repeat sequences, useful
PT   in DNA profiling
XX
PS   Claim 22; Page 86; 111pp; English.
XX
CC   This sequence represents a human DNA marker clone used in the method of
CC   the invention. The method is for isolating a fragment of DNA containing
CC   an intermediate tandem repeat (ITR) sequence using hybridization
CC   selection, and comprises: (a) providing several DNA fragments, at least
CC   one of which contains an ITR sequence, a region of the DNA fragment which
CC   contains at least one repeat unit consisting of a sequence of five, six
CC   or seven bases repeated in tandem at least two times; (b) providing a
CC   stationary support having at least one oligonucleotide associated with
CC   it, where the oligonucleotide includes a sequence of nucleotides which
CC   complementary to a portion of the ITR sequence; and (c) combining the DNA
CC   fragments with the support under conditions where the DNA fragments
CC   including the DNA fragment containing the ITR sequence hybridize to the
CC   support. The method is particularly used to isolate DNA containing
CC   pentanucleotide tandem repeat sequences as well as to detect target ITR
CC   DNA sequences having a low incidence of stutter artifacts (no more than
CC   2.4%). The method is useful in DNA profiling for linkage analysis,
CC   criminal justice, paternity testing and other forensic and medical uses.
CC   DNA typing is also useful for confirming the lineage of horses, dogs and
CC   other prize animals. The invention overcomes problems related to the use
CC   of microsatellite loci in DNA profiling. The method can detect
CC   polymorphisms with a low incidence of stutter artifacts, which has
CC   previously been a problem in interpreting allelic content of loci. The
CC   development of markers based on larger repeat units, enables easier
CC   separation of the fragments on electrophoretic gels. This allows the
CC   simultaneous analysis of more loci.
XX
SQ   Sequence 1366 BP; 436 A; 56 C; 594 G; 280 T; 0 other;

Query Match             2.5%;  Score 35;  DB 20;  Length 1366;
Best Local Similarity   49.2%;
Matches   92;  Conservative    0;  Mismatches  95;  Indels    0;  Gaps    0;

Qy       997 gagcagcatggagtcaaggctgtgtttatgaataaacatgactggtatgaagaagaatatcca 1056
             ||| ||| |||| | |  | ||| | |  |||  |  |||  |||| |||  ||| ||| |  ||
Db       988 gagtagaacggagtgattggggtggaatggaatagagtggaatggaatggaatgagtggagtgg 1047

Qy      1057 actacacctagctcttttgtttaaatggcttatgtattcttaattgggctgtatgac 1116
             | || ||  |  | |||| | | |||||  ||| |||| |||||| ||| ||||| ||
Db      1048 agtagaacggagtgattggaatggaatacagtagagtggaatggcagtggagtggagtgga 1107

Qy      1117 ttaaagaaactgcagggaaacctcgggaagagaacgcaggtccttgtctatgaggcgtggc 1176
             || |||  | || || | |||  |  ||||| | | || |  |  ||| |||| || |||||
Db      1108 atggaatggaagtggagtggcatggaaaggaatggagaggaatggaatggaatgga 1167

Qy      1177 atggaat 1183
             |||||||
Db      1168 atggaat 1174
```

```
RESULT  4
XX
ID   X84332 standard; DNA; 5059 BP.
XX
AC   X84332;
XX
DT   08-SEP-1999 (first entry)
XX
DE   Stealth virus nucleic acid clone, SEQ ID NO: 24.
XX
KW   Stealth virus; detection; diagnosis; infection; ss.
XX
OS   Stealth virus.
XX
FH   Key            Location/Qualifiers
FH
FT   misc_difference 3605
FT                  /tag= a
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3610
FT                  /tag= b
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3615
FT                  /tag= c
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3630
FT                  /tag= d
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3631
FT                  /tag= e
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3638
FT                  /tag= f
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3641
FT                  /tag= g
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3642
FT                  /tag= h
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3657
FT                  /tag= i
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3659
FT                  /tag= j
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3681
FT                  /tag= k
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3698
FT                  /tag= l
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"

FT   misc_difference 3703
FT                  /tag= m
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3708
FT                  /tag= n
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3713
FT                  /tag= o
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3720
FT                  /tag= p
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3732
FT                  /tag= q
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3733
FT                  /tag= r
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3740
FT                  /tag= s
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3747
FT                  /tag= t
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3750
FT                  /tag= u
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3751
FT                  /tag= v
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3766
FT                  /tag= w
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3778
FT                  /tag= x
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3792
FT                  /tag= y
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3799
FT                  /tag= z
FT                  /note= "this nucleotide is represented as a * in the
FT                  specification, and is included to maintain the
FT                  base numbering given in the specification"
FT   misc_difference 3800
```

```
FT   misc_difference  3812
FT                    /*tag=  aa
FT                    /note=  "this nucleotide is represented as a * in the
FT                    specification, and is included to maintain the
FT                    base numbering given in the specification"
FT   misc_difference  3813
FT                    /*tag=  ab
FT                    /note=  "this nucleotide is represented as a * in the
FT                    specification, and is included to maintain the
FT                    base numbering given in the specification"
FT   misc_difference  3818
FT                    /*tag=  ac
FT                    /note=  "this nucleotide is represented as a * in the
FT                    specification, and is included to maintain the
FT                    base numbering given in the specification"
FT   misc_difference  3819
FT                    /*tag=  ad
FT                    /note=  "this nucleotide is represented as a * in the
FT                    specification, and is included to maintain the
FT                    base numbering given in the specification"
FT   misc_difference  3824
FT                    /*tag=  ae
FT                    /note=  "this nucleotide is represented as a * in the
FT                    specification, and is included to maintain the
FT                    base numbering given in the specification"
FT   misc_difference  3827
FT                    /*tag=  af
FT                    /note=  "this nucleotide is represented as a * in the
FT                    specification, and is included to maintain the
FT                    base numbering given in the specification"
FT   misc_difference  3841
FT                    /*tag=  ag
FT                    /note=  "this nucleotide is represented as a * in the
FT                    specification, and is included to maintain the
FT                    base numbering given in the specification"
FT   misc_difference  3853
FT                    /*tag=  ah
FT                    /note=  "this nucleotide is represented as a * in the
FT                    specification, and is included to maintain the
FT                    base numbering given in the specification"
FT   misc_difference  3857
FT                    /*tag=  ai
FT                    /note=  "this nucleotide is represented as a * in the
FT                    specification, and is included to maintain the
FT                    base numbering given in the specification"
FT   misc_difference  3860
FT                    /*tag=  aj
FT                    /note=  "this nucleotide is represented as a * in the
FT                    specification, and is included to maintain the
FT                    base numbering given in the specification"
FT   misc_difference  3863
FT                    /*tag=  ak
FT                    /note=  "this nucleotide is represented as a * in the
FT                    specification, and is included to maintain the
FT                    base numbering given in the specification"
FT                    /*tag=  al
```

```
Query Match     2.5%;  Score 35;  DB 20;  Length 5059;
Best Local Similarity   15.7%;  Pred. No. 3;
Matches   53;  Conservative  120;  Mismatches  164;  Indels  0;  Gaps  0;

Qy      166 caaggttattctaccaattccagattgcacagtatgggttaagtcactacagcaagaat 225
            ::|:   |:|| ::|  :::|| ||:| :|:  |:||  :||||| ||  ||:   :|||
Db     2299 srmthcrrsndnghcmvrhcmvrscmvgnmaanamsrydvddwthcncnithscasandnsvrat 2358

Qy      226 ctaactgaaaaacccctcatatagaggtatatgaacagcaagaacgagaaggacaaaaac 285
            :|::   |:|:|| |::::|:|: |:::| |::|:| :| |:||::::|:|:||| |::|::
Db     2359 hramsthnsrtdsncswdbctdtdsrtthradtrmasngnradngramnrvrmnrdrncs 2418

Qy      286 agcaagcccaatgacctgtgtgccaaggcctgctttatggctagtgtggctgataag 345
            |:||:|| |:| |:|::|:|  :|||   :||| ||::|||||||||  :| ||| :|
Db     2419 andccasnamarsncdrncswrrsntbtwndrntcnsthatcarymatchdtsmarrgns 2478
```

```
C69113

Qy      346 tcaagattcaccaatgttaaacagttcattgctccagaaaccagtgaaggtgtatccttg 405
            |::|    |:|| |:| |||| ||||:| ||| |:| ||| :|||||||| || |:|:|
Db     2479 thhcmvgnmamsarzshwndscssnthgnmthysncdadstranhcmvcmrssnctdsbas 2538

Qy      406 caactgggaacacaaaagattttatttcattgacctcaagttcctcaacaaatgga 465
            |::|    |:|||:|:| |   :||| |:| |:|: ||||:| ||:|| :| || ||
Db     2539 arsandtntaycdsratastrtnsrshcncasatshcmventansmaddtnartncdings 2598

Qy      466 agcgtgtctgtggttctggagacagacagaaaagaatc 502
            ||:||::||||:||:||:| :|:| |:|:||| ::|:| |
Db     2599 ncsntrsntnthabratryadatcdadstranmtdsnc 2635

ID   C69113 standard; DNA; 2248 BP.
XX
AC   C69113;
XX
DT   31-JAN-2001 (first entry)
XX
DE   Human secreted protein gene 9 clone HYEOF33.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; ss.
XX
OS   Homo sapiens.
XX
PN   WO200005371-A1.
XX
PD   21-SEP-2000.
XX
PF   16-MAR-2000; 2000WO-US06783.
XX
PR   18-MAR-1999;  99US-0125055.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Ni J, Ebner R, Rosen CA, Shi Y, Birse C, Florence K;
PI   Komatsoulis G, Lafleur DW, Moore PA, Olsen HS, Young PE;
XX
DR   WPI; 2000-594448/56.
DR   P-PSDB; B38013.
XX
PT   New nucleic acid molecules encoding 27 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives -
XX
PS   Claim 1; Page 387; 453pp; English.
XX
CC   The invention relate to the isolation of genes C69084-C69119 encoding
CC   27 human secreted proteins B37984-B38019. The genes can be used to
CC   generate fusion proteins by linking to the gene for the human
CC   immunoglobulin G Fc portion (C69075) for increasing the stability of
CC   the fusion protein as compared to the human protein only. The genes and
CC   proteins are useful for preventing, ameliorating or treating medical
CC   conditions, e.g. by protein or gene therapy. The genes are isolated
CC   from a range of human tissues disclosed in the specification. The
CC   nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC   (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC   healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC   and (f) infectious diseases such as viral, bacterial, fungal and
CC   parasitic infections.
XX
```

SQ   Sequence 2248 BP; 657 A; 460 C; 518 G; 609 T; 4 other;

Qy   Query Match           2.5%;                Score 34.4;  DB 21;  Length 2248;
     Best Local Similarity 49.4%;               Pred. NO. 2.9;
     Matches 89;  Conservative 0;  Mismatches 91;  Indels 0;  Gaps 0;

Qy    418 acaaaagaatttattcattcattgacctcaagttcttaacaaatgggaagcgtgtctgtg 477
          |||||  ||||  || ||| || ||  |   |||  | |||    |  |  |||||| |||
Db   1826 acatgtgatttttaaaatcatttgactactgggtgtataaatgtgtttgtgtctgtatg 1885

Qy    478 gttctgagacgacagaaagagaaatcagctcttcactgtacattatgtctcaaataccag 537
          | || |   |  |     || |  |||  | || |   ||| ||||||| |||||   |
Db   1886 tatttatagatgtttcttaaggtacctgaaaactcatcattatgtatgccacataatg 1945

Qy    538 ctaattgctcttaaagaagagagacatatactatggcatcgggcccagaacatcatggagc 597
          | ||| | |  ||  |  || | || ||| |||||| || ||  ||||   |||||||||||
Db   1946 acatttcagtcagtggtagactacacatatatgtatggttccaataagaattatatggagc 2005

RESULT 6
X20393/c

ID   X20393 standard; DNA; 571 BP.
XX
AC   X20393;
XX
DT   04-MAY-1999 (first entry)
XX
DE   Borrelia burgdorferi polynucleotide sequence #146.
XX
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9858943-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998; 98WO-US12764.
XX
PR   03-SEP-1997;  97US-0057483.
PR   20-JUN-1997;  97US-0050359.
PR   22-JUL-1997;  97US-0053344.
PR   22-JUL-1997;  97US-0053377.
XX
PS   Claim 1; Page 1109; 1128pp; English.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
XX
DR   WPI; 1999-081217/07.
XX
PI   White OR;
PI   Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
CC   X20248 to X20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
XX
SQ   Sequence 571 BP; 198 A; 76 C; 54 G; 241 T; 2 other;

RESULT 7
X20260

ID   X20260 standard; DNA; 9542 BP.
XX
AC   X20260;
XX
DT   04-MAY-1999 (first entry)
XX
DE   Borrelia burgdorferi polynucleotide sequence #13.
XX
KW   Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW   epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW   infection; diagnosis; characterisation; detection; ds.
XX
OS   Borrelia burgdorferi.
XX
PN   WO9858943-A1.
XX
PD   30-DEC-1998.
XX
PF   18-JUN-1998; 98WO-US12764.
XX
PR   03-SEP-1997;  97US-0057483.
PR   20-JUN-1997;  97US-0050359.
PR   22-JUL-1997;  97US-0053344.
PR   22-JUL-1997;  97US-0053377.
XX
PS   Claim 1; Page 920-925; 1128pp; English.
XX
PT   New isolated Borrelia burgdorferi nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of infections, particularly Lyme disease
XX
DR   WPI; 1999-081217/07.
XX
PI   White OR;
PI   Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (MEDI-) MEDIMMUNE INC.
XX
CC   X20248 to X20402 represent polynucleotide sequences isolated from
CC   Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC   the detection, diagnosis, characterisation, prevention and therapy of
CC   Bb infections, e.g. Lyme disease. They can also be used for the
CC   production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC   to a family of motile, spiral-shaped bacteria called Spirochetes.
CC   Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC   endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC   Lyme disease.
XX
SQ   Sequence 9542 BP; 3812 A; 1160 C; 1113 G; 3457 T; 0 other;

Qy   Query Match           2.4%;                Score 33.8;  DB 20;  Length 571;
     Best Local Similarity 49.7%;               Pred. NO. 1.9;
     Matches 86;  Conservative 0;  Mismatches 87;  Indels 0;  Gaps 0;

Qy    403 ttgcaactgggggaacacaaaagatttattcattcattgacctcaagttcttaacaaat 462
          ||| ||  |     || |   | || ||| |  || || || |  ||||||| |||| |
Db    504 TTGTAAATAATTAGAGAATAGGTTTTTATAAGTTCTTTAAAATAAGATTTTATAAAAAG 445

Qy    463 ggaagcgtgtctgtggttctggagacgacagaaagagaatcagctcttcactgtacattat 522
          || |   ||  | || | |  |   | | |  ||||| || |  | || || | | |||
Db    444 CTTTATATANTTTTGTANTTTTTTATAGACCGCAGTGTAATAATATTAATTAATTGATTTTAATTAA 385

Qy    523 gtctcaaataccagctaattgctcttaaagaagagagacatatactatggcat 575
          | | |||||| |  ||||| | |  || | | ||  || ||| |||| |||
Db    384 GGTTGAACTAAACTAAAATATATAGTTTTGTTAGGAAAATAATTTTTCATTATTTTCCT 332

Query Match           2.4%;                Score 33.8;  DB 20;  Length 9542;

Best Local Similarity 49.7%; Pred. No. 11; Mismatches 87; Indels 0; Gaps 0;
Matches 86; Conservative 0;

Qy  403  ttgcaactgggaaacacaaagattttattattcattgacctcaagttcttaacaat 462
         | ||  || ||| || ||||| |||||| ||||| ||| ||| ||| || ||
Db  8979 ttgtaataattagagatagtttttataagttctttaaaataagatttataaag 9038

Qy  463  ggaagcgtgtctgtggttctcggagacgacagaaaagaatcagctcttcactgtacattat 522
         || | |||  || |||| || || ||| |||||| |||| ||  |||||  ||| ||
Db  9039 ctttatatttttgtatttttatagaccgcagtgtaataattgatttaattaa 9098

Qy  523  gtctcaataccacagctaattcgtcttttaaagaaagaagagacatatactatggcat 575
         | ||| ||||| ||  ||||| ||| ||||| | |||| ||||| || ||| |||
Db  9099 ggttgaactaaactaaaatatagtttgtaggaataataattctcattatttcct 9151

RESULT  8
ID    C94048/c
ID    C94048 standard; cDNA; 283 BP.
AC    C94048;
XX
DT    19-FEB-2001 (first entry)
XX
DE    Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:543.
XX
KW    Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
KW    flea infestation; vaccine; antiparasitic; therapeutic target;
KW    diagnosis; detection; ss.
XX
OS    Ctenocephalides felis.
XX
PN    WO200061621-A2.
XX
PD    19-OCT-2000.
XX
PF    07-APR-2000; 2000WO-US09437.
XX
PR    09-APR-1999;    99US-0128704.
XX
PA    (HESK-) HESKA CORP.
XX
PI    Brandt KS, Gaines PJ, Stinchcomb DT, Wisnewski N;
XX
DR    WPI; 2000-656323/63.
XX
PT    Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT    acids useful for the prevention, diagnosis and treatment of flea
PT    infestations -
XX
PS    Claim 26; Page 425; 964pp; English.

CC    The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC    acids which are expressed in hindgut and Malpighian tubule.(HMT) tissue
CC    or head and nerve cord (HNC) tissue. The invention also relates to the
CC    encoded proteins. The invention additionally encompasses expression
CC    constructs, recombinant viruses and recombinant cells comprising the
CC    nucleic acids of the invention, recombinant production of the proteins,
CC    antibodies against the proteins, a method of identifying inhibitors of
CC    the proteins, and compositions comprising the inhibitors for
CC    administration to an animal. The nucleic acids, and the proteins they
CC    encode may be used in the prevention, treatment and diagnosis of diseases
CC    associated with flea infestations. For example, the nucleic acids may be
CC    used to produce an HMT or HNC protein according to standard recombinant
CC    DNA methodology by inserting the nucleic acids into a host cell and
CC    culturing the cell to express the protein. The HMT and HNC nucleic acids
CC    may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC    and quantitate the presence of cat flea or other homologous nucleic acid
CC    sequences in samples. They may also be used to study the expression and
CC    function of the proteins and their role in metabolism. The HMT and HNC
CC    proteins may be used as antigens in the production of specific
CC    antibodies, and in assays to identify modulators (agonists and
CC    antagonists) of HMT and/or HNC protein expression and activity. The

CC    anti-HMT/HNC protein antibodies and antagonists may also be used to
CC    downregulate protein expression and activity. The antibodies may also be
CC    used as diagnostic agents for detecting the presence of flea polypeptides
CC    in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC    present sequence represents a cat flea HMT cDNA of the invention.
XX
SQ    Sequence 283 BP; 78 A; 73 C; 47 G; 84 T; 1 other;

Qy  376  gtccagaaaccagtgaagtgtatcctgcaactgggggaacaacaaaagagattctatt 435
         | | |||| ||||  || ||||| ||| ||  ||| ||| |||||||| ||||| |||
Db  159  GTTTCAGAATTGGGAGGAGAAGCTGGAACCGTCCAAATTTAGAGACTTGAATGTTGATTGTTAAT 100

RESULT  9
ID    C39833
ID    C39833 standard; DNA; 1231 BP.
AC    C39833;
XX
DT    17-OCT-2000 (first entry)
XX
DE    Arabidopsis thaliana DNA fragment SEQ ID NO: 26071.
XX
KW    Hybridisation assay; genetic mapping; gene expression control;
KW    protein identification; signal transduction pathway;
KW    metabolic pathway; promoter; termination sequence; ss.
XX
OS    Arabidopsis thaliana.
XX
PN    EP1033405-A2.
XX
PD    06-SEP-2000.
XX
PF    25-FEB-2000; 2000EP-0301439.
XX
PR    25-FEB-2000;    2000EP-0301439.
PR    05-MAR-1999;    99US-0121825.
PR    09-MAR-1999;    99US-0123180.
PR    16-APR-1999;    99US-0128714.
PR    19-APR-1999;    99US-0128845.
PR    21-APR-1999;    99US-0123548.
PR    23-APR-1999;    99US-0125788.
PR    23-APR-1999;    99US-0130449.
PR    29-MAR-1999;    99US-0126264.
PR    01-APR-1999;    99US-0126785.
PR    06-APR-1999;    99US-0130510.
PR    28-APR-1999;    99US-0130891.
PR    30-APR-1999;    99US-0127462.
PR    30-APR-1999;    99US-0131449.
PR    04-MAY-1999;    99US-0132048.
PR    05-MAY-1999;    99US-0132407.
PR    06-MAY-1999;    99US-0132484.
PR    06-MAY-1999;    99US-0132485.
PR    07-MAY-1999;    99US-0132886.
PR    11-MAY-1999;    99US-0132487.
PR    14-MAY-1999;    99US-0132863.
PR    14-MAY-1999;    99US-0134256.
PR    14-MAY-1999;    99US-0134218.
PR    14-MAY-1999;    99US-0134219.
PR    14-MAY-1999;    99US-0134221.
PR    18-MAY-1999;    99US-0134370.
PR    19-MAY-1999;    99US-0134768.
PR    19-MAY-1999;    99US-0134941.

Query Match   2.48;   Score 33.6;  DB 21;   Length 283;
Best Local Similarity  56.2%;  Pred. No. 1.4;
Matches  63;  Conservative 0;  Mismatches 49;  Indels 0;  Gaps 0;

Qy  436  tcattgacctcaagttcttaacaaatggaagccgtgtctgtggttctgggga 487
         | | |||||||| ||| | ||| |||| || ||| ||||| |||| || |||
Db   99  TCATTTCAACGTAAATTCGTCAACGAAGTCAGGCGTTGTGATGAGATGGAGA 48

TCATTTCAAACGTAAAATTCGTCAACGAAGTCAGCCGTTGTGATGAGATGTTAAT 100

```
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139494.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 24-JUN-1999; 99US-0140353.
PR 28-JUN-1999; 99US-0140354.
PR 29-JUN-1999; 99US-0140695.
PR 30-JUN-1999; 99US-0140823.
PR 01-JUL-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 02-JUL-1999; 99US-0141842.
PR 06-JUL-1999; 99US-0142154.
PR 08-JUL-1999; 99US-0142055.
PR 09-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144333.
PR 21-JUL-1999; 99US-0144334.
PR 21-JUL-1999; 99US-0144335.
PR 22-JUL-1999; 99US-0144352.
PR 22-JUL-1999; 99US-0144632.
PR 22-JUL-1999; 99US-0144814.
PR 22-JUL-1999; 99US-0144884.
PR 23-JUL-1999; 99US-0145086.
PR 23-JUL-1999; 99US-0145085.
PR 23-JUL-1999; 99US-0145087.
PR 26-JUL-1999; 99US-0145089.
PR 27-JUL-1999; 99US-0145192.
PR 27-JUL-1999; 99US-0145145.
PR 28-JUL-1999; 99US-0145218.
PR 28-JUL-1999; 99US-0145224.
PR 28-JUL-1999; 99US-0145276.
PR 28-JUL-1999; 99US-0145913.
PR 28-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0155659.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 12-OCT-1999; 99US-0158029.
PR 13-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 14-OCT-1999; 99US-0159293.
PR 14-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 18-OCT-1999; 99US-0159329.
PR 21-OCT-1999; 99US-0159330.
PR 21-OCT-1999; 99US-0159331.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0159637.
PR 22-OCT-1999; 99US-0159638.
PR 22-OCT-1999; 99US-0160741.
PR 22-OCT-1999; 99US-0160767.
PR 25-OCT-1999; 99US-0160768.
PR 25-OCT-1999; 99US-0160770.
PR 25-OCT-1999; 99US-0160814.
PR 26-OCT-1999; 99US-0160815.
PR 26-OCT-1999; 99US-0160980.
PR 26-OCT-1999; 99US-0160981.
PR 26-OCT-1999; 99US-0161406.
PR 27-OCT-1999; 99US-0161405.
PR 28-OCT-1999; 99US-0161404.
PR 28-OCT-1999; 99US-0161359.
PR 28-OCT-1999; 99US-0161360.
```

```
PR   26-OCT-1999;    99US-0161361.
PR   28-OCT-1999;    99US-0161920.
PR   28-OCT-1999;    99US-0161992.
PR   28-OCT-1999;    99US-0161993.
PR   29-OCT-1999;    99US-0162142.
XX

Query Match             2.4%;    Score 33.6;  DB 21;  Length 1231;
Best Local Similarity  51.3%;    Pred. No. 3.5;
  Matches   78;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy   1031 aacatgactggtgatgaagaataatatccaactacaccactagctctttgttttaaatggctta 1090
          ||||||  |||| |||    |||| || |||  |||||||| | ||| |||||||  |||
Db    454 aaaaatgtgtggaataggcaaaatgcaaatctcctattgatcttcgggacaaaatgtggta 513

Qy   1091 tgtatcctttaattgggctgtatgactaaaagaaaactcgcaggggaaaactcggggaaag 1150
          ||| ||||||||  |||||||| || || |||| ||||  ||||| |||||||||  |||
Db    514 gttagtaataaaattcgattcgttcctagccagtatctgccttctcaataccaccattaag 573

Qy   1151 aagcgaggtccttgtatggagcgtggacatggaa 1182
          || |||||||  |||||||| || ||| |||||| ||
Db    574 aacagaggtcatgatatcatgttgaaattcaa 605
```

## RESULT 10

```
ID   V44080/c
    V44080 standard; cDNA; 1392 BP.
XX
AC   V44080;
XX
DT   01-OCT-1998  (first entry)
XX
DE   mu-NEO-r DNA fragment #2.
XX
KW   DHFR; NEO-r; expression vector; neomycin phosphotransferase gene;
KW   expression-inducing promoter; G418-resistant cell; ds.
XX
OS   Unknown.
XX
PN   JP1017169169-A.
XX
PD   07-JUL-1998.
XX
PF   25-AUG-1997;    97JP-0243357.
XX
PR   23-AUG-1996;    96JP-0222884.
XX
PA   (IMMO ) IMMUNO JAPAN INC.
XX
DR   WPI; 1998-430954/37.
XX
PT   Expression vector for mammalian cell - comprises phosphotransferase
PT   gene cistron and multi-cloning site for gene recombination
XX
PS   Disclosure; Page 25-26; 35pp; Japanese.
XX
CC   V44053-V44087 are DNA sequences used in a method resulting in the
CC   construction of a novel expression vector. This vector contains a
CC   neomycin phosphotransferase gene cistron and a multicloning site for
CC   gene recombination having a strong expression-inducing promoter in the
CC   plasmid and induces high productivity of recombinant protein in a
CC   mammalian host cell. Also claimed is a preparation of a high
CC   productivity transformant of a protein which is a G418-resistant cell
CC   clone which involves transforming a host cell with the expression vector
CC   recombined with a protein gene. This expression vector can produce high
CC   levels of the recombinant protein.
XX
SQ   Sequence 1392 BP; 348 A; 370 C; 426 G; 248 T; 0 other;

Query Match             2.4%;    Score 33.6;  DB 19;  Length 1392;
Best Local Similarity  55.0%;    Pred. No. 3.8;
  Matches   66;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;
```

```
Qy   1175 gcatggaatcccctaaagccatgtctcccttgtacgacaactggctcaggagaaccatctatg 1234
          ||||| || ||| |||  || ||| ||||||| ||||||| ||| ||  |||| |||| |||
Db    764 GCATTGACCTGTCGCAAGCCCACTCTCTCCCCTTTGCTCCTGTGTTTCTCCAGGATCACCTCTG 705

Qy   1235 acctccggcacttcatgtcttgggcattgccccaacctgtccccgctgggactataccacaca 1294
          |||||| || || |||| | ||  |||| ||||||||  |||| || |||| |||||  ||
Db    704 TCCCCCTTCAGTCCTGTCGTTGGGGGGCGGCCCTGGCACCTGGAAGGGCCCTGTGGACCTATTGCCTCCA 645
```

## RESULT 11

```
ID   Q92554/c
    Q92554 standard; cDNA; 1534 BP.
XX
AC   Q92554;
XX
DT   10-FEB-1996  (first entry)
XX
DE   Bovine conglutinin nucleic acid.
XX
KW   Conglutinin; therapeutic; diagnostic; ss.
XX
OS   Bos taurus.
XX
FH   Key              Location/Qualifiers
FH
FT   CDS              231..1346
FT                    /*tag= a
FT                    /note= "bovine conglutinin"
FT   mat_peptide      291..1343
FT                    /*tag= b
FT   sig_peptide      231..290
FT                    /*tag= c
XX
PN   WO9516697-A1.
XX
PF   22-JUN-1995.
XX
PF   14-DEC-1994;    94WO-US14656.
XX
PR   16-DEC-1993;    93US-0168458.
XX
PA   (IMMU-) APPLIED IMMUNE SCI INC.
PA   (UYBO-) UNIV BOSTON.
XX
PI   Lee YM, Leiby KR, Okarma TB, Sastry K;
XX
DR   WPI; 1995-233510/30.
DR   P-PSDB; R75642.
XX
PT   New nucleic acid encoding bovine conglutinin and its fragments - and
PT   related vectors, host cell, etc. useful e.g. in modulating
PT   conglutinin expression or for selective removal of cpds., carrying
PT   particular carbohydrate residues.
XX
PS   Claim 1; Fig 1A-1B; 29pp; English.
XX
CC   The nucleic acid is used in the production of bovine conglutinin,
CC   either the complete protein, or amino acid regions 1-351, 197-351 or
CC   250-351. Oligomers based on the nucleic acid are used to modulate
CC   conglutinin production in livier cells e.g. to modulate the immune
CC   response. They are also useful as primers and probes for
CC   therapeutic or diagnostic use. Conglutinin, or its carbohydrate-
CC   binding fragments, are used to bind cells carrying the C1q receptor.
CC   Solid supports are used to remove compounds (especially proteins or
CC   peptides, e.g. immune complexes) having a carbohydrate component able
CC   to bind to the lectin region of conglutinin, optionally with recovery
CC   of bound compounds.
XX
SQ   Sequence 1534 BP; 395 A; 398 C; 458 G; 283 T; 0 other;

Query Match             2.4%;    Score 33.6;  DB 16;  Length 1534;
Best Local Similarity  55.0%;    Pred. No. 4.1;
  Matches   66;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;
```

Qy   1175 gcatggaatcccttaaagccatgctccccttgtacgacaactggctcaggaaggaaccactctatg 1234
          ||| ||||||||||||||| || ||||||||||| ||||| || ||||||||||| ||
Db    889 GCATTGACCTCTGCAAGCCACTGCTCCCCCTTTGCCTCCTGTTTTCTCCAGGATCACCTCTG 830

Qy   1235 aacctccggcacttcatgcttgcccccaacttggcccgctgggacttatcacaca 1294
          ||| |||||||||||||  |||||||| ||||||||  ||||||| |||
Db    829 TCCCCCTTCAGTCCTGGGGGGCCCCCTGGGCACCTGAAGGGCCCTTGTGGACCTATGGCTCCA 770

RESULT 12
ID   X60775/c
AC   X60775;
XX
XX
XX
DT   20-JUL-1999  (first entry)
XX
DE   Soybean cytochrome P450 enzyme, CYP77A3 encoding cDNA.
XX
KW   Soybean; Solanaceae crop plant; cytochrome P450; transgenic plant;
KW   enzyme; phenylurea herbicide; herbicide resistance; ss.
XX
OS   Glycine max.
XX
XX
PN   WO9919493-A2.
XX
PD   22-APR-1999.
XX
PF   05-OCT-1998;
PR   10-OCT-1997;  97US-0948564.
XX
DR   WPI; 1999-302532/25.
DR   P-PSDB; Y09185.
XX
PT   DNA encoding soybean cytochrome P450 enzymes
XX
PS   Claim 1; Page 48-51; 93pp; English.
XX
CC   The invention provides new DNA molecules (X60773-X60781) encoding soybean
CC   cytochrome P450 enzymes (Y09183-Y09191) respectively. The DNA encoding
CC   the cytochrome P450 enzymes is useful for transformation of Solanaceae
CC   crop plants. Transgenic plants comprising DNA constructs having the P450
CC   encoding nucleic acid sequences are resistant to phenylurea herbicides.
CC   The transgenic plants have increased resistance to phenylurea herbicides
CC   compared to wild-type plants of the same species. The plant crops, e.g.
CC   turfgrass, tobacco, potato, corn, rice, cotton, soybean, rape,
CC   wheat, oats, barley or rice are particularly resistant to fluometuron,
CC   linuron, chlortoluron or diuron.
XX
SQ   Sequence 1644 BP; 482 C; 366 A; 400 G; 396 T; 0 other;

Query Match       2.4%;  Score 33;  DB 20;  Length 1644;
Best Local Similarity   55.8%;  Pred. No. 6.5;
  Matches  63;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    399 atccttgcaactggggaacacaaagatttattcattgacctcaagttcttaac 458
          |  || |||||| |||| ||||||||  || |  |||| |||| ||||| ||||
Db    756 AACCTGTTCTGTCGAACCTCCAAGGCTTTCTTTTTCTTTGCTTTGAGAAAAGGGGCTTAG 697

Qy    459 aaaatggaagtgtctgtggttctggagacacagaaagaaatcagtcttca 511
          ||||||||||| ||||||||| |||| || |||| ||||||| ||||| ||
Db    696 AATTGGAGATAGTCATCAATTCTCGGGTCCAAAGTGATGAGAACACTCTTCA 644

RESULT 13
A97911

---

ID   A97911 standard; DNA; 3970 BP.
XX
AC   A97911;
XX
DT   19-JAN-2001  (first entry)
XX
DE   L. mesenteroides alternan sucrase DNA from clone AS-19B2.
XX
KW   Alternan sucrase; glucosyltransferase; fructose; cosmetic; foodstuff;
KW   syrup; ds.
XX
OS   Leuconostoc mesenteroides.
XX
XX
DR   .DE1905069-A1.
XX
PD   10-AUG-2000.
XX
PF   08-FEB-1999;  99DE-1005069.
PR   08-FEB-1999;  99DE-1005069.
XX
PI   Kossmann J, Welsh T, Quanz M, Knuth K;
XX
PA   (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PT   New nucleic acid encoding recombinant Leuconostoc mesenteroides
PT   alternan sucrase protein and methods of alternan and fructose
PT   production
XX
PS   Example 1; Page 41-43; 64pp; German.
XX
CC   This invention describes a novel nucleic acid molecule (I) encoding an
CC   alternan sucrase (E.C. 2.4.1.140 - an enzyme, that belongs to the
CC   glucosyltransferase group) The recombinant, purified alternan sucrase
CC   gene is useful for the fermentative production of alternan (a
CC   carbohydrate) and/or fructose by secreting the enzyme into a
CC   saccharose-containing culture medium. Alternatively, the enzyme is
CC   contacted with a saccharose-containing solution. The alternan and/or
CC   fructose is then isolated from the medium. Cosmetic products or
CC   foodstuffs containing alternan can be produced. Recombinant production of
CC   alternan sucrase is advantageous as it provides a cost effective means of
CC   producing fructose for high fructose containing syrups, production of
CC   which previously has been achieved by costly production from maize
CC   starch. This sequence encodes a Leuconostoc mesenteroides alternan
CC   sucrase protein isolated from clone AS-19B2 which is described in the
CC   method of the invention.
XX
SQ   Sequence 3970 BP; 1354 A; 651 C; 823 G; 1142 T; 0 other;

Query Match       2.3%;  Score 32.6;  DB 21;  Length 3970;
Best Local Similarity   52.6%;  Pred. No. 15;
  Matches  71;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy    426 ttttattatttcattgacctcaagttcttaacaaatggaagcgtgtctgttctgga 485
          ||| || |||||| ||||||| |||  ||  ||  |||||| || ||| || || |||
Db   1605 tgttggtaataaccctgacttaagttgaacaatgataagacacattaccttgcatatgg 1664

Qy    486 gacgacagaaagaaatcagtcttcactgtacattatgtcctcaaataccagctaattgc 545
          ||| || ||||||||| ||| || ||  || |  || |||||||||  |||||| | ||
Db   1665 aaaggcgcataagacatcaacttaccgtgccttagtattatcaaatgactcaggaattga 1724

Qy    546 ttttaaagaaagaga 560
          || |  || |||| ||
Db   1725 tgtttatgatatga 1739

RESULT 14
A97917

ID   A97917 standard; DNA; 6204 BP.

## Left record

```
XX
AC   A97917;
XX
DT   19-JAN-2001  (first entry)
XX
DE   L. mesenteroides alternan sucrase insertion sequence DNA.
XX
KW   Alternan sucrase; glucosyltransferase; fructose; cosmetic; foodstuff;
KW   syrup; ds.
XX
OS   Leuconostoc mesenteroides.
XX
PN   DE19905069-A1.
XX
PD   10-AUG-2000.
XX
PF   08-FEB-1999;   99DE-1005069.
XX
PR   08-FEB-1999;   99DE-1005069.
XX
PA   (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Kossmann J,  Welsh T,  Quanz M,  Knuth K;
XX
DR   WPI; 2000-550294/51.
XX
PT   New nucleic acid encoding recombinant Leuconostoc mesenteroides
PT   alternan sucrase protein and methods of alternan and fructose
PT   production .
XX
PS   Example 4; Page 51-54; 64pp; German.
XX
CC   This invention describes a novel nucleic acid molecule (I) encoding an
CC   alternan sucrase (E.C. 2.4.1.140 - an enzyme, that belongs to the
CC   glucosyltransferase group) The recombinant, purified alternan sucrase
CC   gene is useful for the fermentative production of alternan (a
CC   carbohydrate) and/or fructose by secreting the enzyme into a
CC   saccharose-containing culture medium. Alternatively, the enzyme is
CC   contacted with a saccharose-containing solution. The alternan and/or
CC   fructose is then isolated from the medium. Cosmetic products or
CC   foodstuffs containing alternan can be produced. Recombinant production of
CC   alternan sucrase is advantageous as it provides a cost effective means of
CC   producing fructose for high fructose containing syrups, production of
CC   which previously has been achieved by costly production from maize
CC   starch. This sequence represents a Leuconostoc mesenteroides alternan
CC   sucrase protein encoding DNA insertion sequence which is described in the
CC   method of the invention.
XX
SQ   Sequence 6204 BP; 2101 A; 1038 C; 1354 G; 1711 T; 0 other;

Query Match             2.3%;
Best Local Similarity   52.6%;
  Matches   71;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;
  Score 32.6;   DB 21;   Length 6204;
  Pred. No. 20;

Qy    426 tttattatttcattttgacctcaagttcttaacaaatggaagcgtgtctgtggtttctgga 485
          |||  ||| ||||  |||| | ||||  |||  ||  ||  |  ||||  ||  ||  | |||
Db   2787 tgttggtaataacctgaccttaagttgaacaatgatataagacacattaccttgcatatggg 2846

Qy    486 gacgacagaaaagaatcagctcttcacttgtacacagctaccaagctaattgc 545
          ||| |  | ||||||||| ||  | |||| ||||| ||||  ||
Db   2847 aaaggcgcataagaatcaacttaccgtgccttagtattattatcaaatgactcaggaattga 2906

Qy    546 ttttaaagaaagaga 560
          |||  || |||| ||
Db   2907 tgtttatgatagtga 2921
```

## Right record

```
AC   A97914;
XX
DT   19-JAN-2001  (first entry)
XX
DE   L. mesenteroides alternan sucrase 3'UTR.
XX
KW   Alternan sucrase; glucosyltransferase; fructose; cosmetic; foodstuff;
KW   syrup; ds.
XX
OS   Leuconostoc mesenteroides.
XX
PN   DE19905069-A1.
XX
PD   10-AUG-2000.
XX
PF   08-FEB-1999;   99DE-1005069.
XX
PR   08-FEB-1999;   99DE-1005069.
XX
PA   (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Kossmann J,  Welsh T,  Quanz M,  Knuth K;
XX
DR   WPI; 2000-550294/51.
XX
PT   New nucleic acid encoding recombinant Leuconostoc mesenteroides
PT   alternan sucrase protein and methods of alternan and fructose
PT   production
XX
PS   Example 2; Page 46-50; 64pp; German.
XX
CC   This invention describes a novel nucleic acid molecule (I) encoding an
CC   alternan sucrase (E.C. 2.4.1.140 - an enzyme, that belongs to the
CC   glucosyltransferase group) The recombinant, purified alternan sucrase
CC   gene is useful for the fermentative production of alternan (a
CC   carbohydrate) and/or fructose by secreting the enzyme into a
CC   saccharose-containing culture medium. Alternatively, the enzyme is
CC   contacted with a saccharose-containing solution. The alternan and/or
CC   fructose is then isolated from the medium. Cosmetic products or
CC   foodstuffs containing alternan can be produced. Recombinant production of
CC   alternan sucrase is advantageous as it provides a cost effective means of
CC   producing fructose for high fructose containing syrups, production of
CC   which previously has been achieved by costly production from maize
CC   starch. This sequence represents a Leuconostoc mesenteroides alternan
CC   sucrase 3'-UTR DNA fragment which is described in the method of
CC   the invention.
XX
SQ   Sequence 7387 BP; 2469 A; 1218 C; 1581 G; 2119 T; 0 other;

Query Match             2.3%;
Best Local Similarity   52.6%;
  Matches   71;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;
  Score 32.6;   DB 21;   Length 7387;
  Pred. No. 22;

Qy    426 tttattatttcattttgacctcaagttcttaacaaatggaagcgtgtctgtggtttctgga 485
          |||  ||| ||||  |||| | ||||  |||  ||  ||  |  ||||  ||  ||  | |||
Db   3458 tgttggtaataacctgaccttaagttgaacaatgatataagacacattaccttgcatatggg 3517

Qy    486 gacgacagaaaagaatcagctcttcacttgtacacagctaccaagctaattgc 545
          ||| |  | ||||||||| ||  | |||| ||||| ||||  ||
Db   3518 aaaggcgcataagaatcaacttaccgtgccttagtattattatcaaatgactcaggaattga 3577

Qy    546 ttttaaagaaagaga 560
          |||  || |||| ||
Db   3578 tgtttatgatagtga 3592
```

```
RESULT 15
ID   A97914
XX
AC   A97914 standard; DNA; 7387 BP.
```

Search completed: May 25, 2001, 03:59:24
Job time: 3280 sec

us-09-403-269-12_copy_1_1404.rng

```
                        GenCore version 4.5
                        Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         May 25, 2001, 02:22:24 ; Search time 67.79 Seconds
                                          (without alignments)
                                          3616.200 Million cell updates/sec

Title:          US-09-403-269-12_COPY_1_1404
Perfect score:  1404
Sequence:       1 tccaagctgaattctcatag.........gcagccgggcaaagcacaac 1404

Scoring table:  IDENTITY_NUC
                Gapop 10.0 ; Gapext 1.0

Searched:       302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters:  605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      Issued_Patents_NA:*
                1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 37.4 | 2.7 | 7218 | 1 | US-08-232-463-14 | Sequence 14, Appl |
| 2 | 35.4 | 2.5 | 474 | 4 | US-09-276-531-29 | Sequence 29, Appl |
| 3 | 33 | 2.4 | 1644 | 4 | US-08-948-564-5 | Sequence 5, Appl |
| 4 | 31.8 | 2.3 | 1661 | 1 | US-08-318-492-3 | Sequence 3, Appl |
| 5 | 31.8 | 2.3 | 1661 | 1 | US-08-707-340-3 | Sequence 3, Appl |
| 6 | 31.8 | 2.3 | 1661 | 2 | US-08-994-578-3 | Sequence 3, Appl |
| 7 | 31.4 | 2.2 | 1627 | 1 | US-08-615-170-2 | Sequence 2, Appl |
| 8 | 31.4 | 2.2 | 1666 | 1 | US-08-615-170-4 | Sequence 4, Appl |
| 9 | 31.4 | 2.2 | 51259 | 3 | US-08-781-891-209 | Sequence 209, App |
| 10 | 31 | 2.2 | 3639 | 3 | US-08-737-524B-26 | Sequence 26, Appl |
| 11 | 30.8 | 2.2 | 4041 | 1 | US-08-147-812-4 | Sequence 4, Appl |
| 12 | 30.8 | 2.2 | 4110 | 4 | US-09-123-708-11 | Sequence 11, Appl |
| 13 | 30.8 | 2.2 | 4110 | 4 | US-09-123-624-1 | Sequence 1, Appl |
| 14 | 30.8 | 2.2 | 4165 | 1 | US-08-147-812-6 | Sequence 6, Appl |
| 15 | 30.6 | 2.2 | 1605 | 1 | US-08-701-582D-1 | Sequence 1, Appl |
| 16 | 30.6 | 2.2 | 1710 | 3 | US-09-255-912-1 | Sequence 1, Appl |
| 17 | 30.6 | 2.2 | 1710 | 4 | US-09-082-039A-14 | Sequence 14, Appl |
| 18 | 30.6 | 2.2 | 176373 | 4 | US-09-128-155-17 | Sequence 17, Appl |
| 19 | 30.6 | 2.2 | 1603 | 1 | US-07-885-970A-19 | Sequence 19, Appl |
| 20 | 30.4 | 2.2 | 1603 | 1 | US-08-298-687A-19 | Sequence 19, Appl |
| 21 | 30.4 | 2.2 | 1603 | 1 | US-08-298-829-19 | Sequence 19, Appl |
| 22 | 30 | 2.1 | 3164 | 1 | US-08-188-228-49 | Sequence 49, Appl |
| 23 | 30 | 2.1 | 3164 | 1 | US-08-332-643-43 | Sequence 43, Appl |
| 24 | 30 | 2.1 | 3164 | 1 | US-08-332-638-49 | Sequence 49, Appl |
| 25 | 30 | 2.1 | 4380 | 1 | US-07-582-945-1 | Sequence 1, Appl |
| 26 | 30 | 2.1 | 4380 | 2 | US-08-453-141-1 | Sequence 1, Appl |
| 27 | 30 | 2.1 | 4380 | 4 | US-08-293-314-1 | Sequence 1, Appl |
| 28 | 29.6 | 2.1 | 2924 | 3 | US-09-357-073-1 | Sequence 1, Appll |
| 29 | 29.4 | 2.1 | 1473 | 2 | US-08-502-725-31 | Sequence 31, Appl |
| 30 | 29.4 | 2.1 | 1630 | 2 | US-08-468-856B-2 | Sequence 2, Appll |
| 31 | 29.4 | 2.1 | 1630 | 3 | US-08-468-859A-2 | Sequence 2, Appll |
| 32 | 29.4 | 2.1 | 1886 | 1 | US-07-891-942G-4 | Sequence 4, Appll |
| 33 | 29.4 | 2.1 | 2370 | 2 | US-08-468-856B-6 | Sequence 6, Appll |
| 34 | 29.4 | 2.1 | 2370 | 4 | US-09-838-219B-19 | Sequence 19, Appl |
| 35 | 29.4 | 2.1 | 2370 | 4 | US-09-233-336A-19 | Sequence 19, Appl |
| 36 | 29.4 | 2.1 | 2911 | 2 | US-09-233-752A-19 | Sequence 19, Appl |
| 37 | 29.4 | 2.1 | 2911 | 3 | US-08-588-976-11 | Sequence 11, Appl |
| 38 | 29.4 | 2.1 | 3461 | 3 | US-08-468-859A-6 | Sequence 6, Appll |
| 39 | 29.4 | 2.1 | 3461 | 3 | US-08-588-983-11 | Sequence 11, Appl |
| 40 | 29.4 | 2.1 | 7920 | 4 | US-09-080-983-2 | Sequence 2, Appll |
| 41 | 29.4 | 2.1 | 15500 | 4 | US-09-080-983-1 | Sequence 1, Appll |
| 42 | 29.2 | 2.1 | 2778 | 1 | US-08-202-054-1 | Sequence 1, Appll |
| 43 | 29.2 | 2.1 | 2778 | 1 | US-08-446-923-1 | Sequence 1, Appll |
| 44 | 29 | 2.1 | 423 | 1 | US-09-470-179-176 | Sequence 176, App |
| 45 | 29 | 2.1 | 1834 | 1 | US-08-592-126-90 | Sequence 90, Appl |

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTzgpt-F1s
US-08-232-463-14
```

RESULT 2
US-09-276-531-29/c
Sequence 29, Application US/09276531
Patent No. 6183968
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Reddy, Roopa
APPLICANT: Guegler, Karl J.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/276,531
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/079,677
FILING DATE: March 27, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:

Query Match 2.7%; Score 37.4; DB 1; Length 7218;
Best Local Similarity 6.2%; Pred. No. 0.13;
Matches 26; Conservative 207; Mismatches 188; Indels 0; Gaps 0;

```
Qy     7 ctgaattctcatagctattccaaagtctatacgacagagagcccctttatcacctgatggt  66
           |  ||    ||    ||||   ||||
Db  1485 CTCTAATTCACTATCTATCGCAAGTAGTTAAAGAGATTAGAAGAATTGGTACRRRRRRRR 1426

Qy    67 gtgttatgtctcttgaagctacaatgtggaagtccgagacagagtcaagtgcataagt  126
         RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
Db  1425 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1366

Qy   127 gggtgttgaagtgtacctttatctacacagtggggacctcaaggctattctaccaatc  186
         RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
Db  1365 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1306

Qy   187 cagattgcacagtatgggttaagtcactcacagacaagaaatctaactgaaaacccctcat 246.
         RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
Db  1305 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1246

Qy   247 ataagaggtatatgaaacagcagaagacagggacaaaaacagcaagcccaatgcact  306
         RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
Db  1245 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1186

Qy   307 gtgcccaaggggctgcttatggctagtgtggctgataagtcaagattcaccaatgttaa  366
         RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
Db  1185 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1126

Qy   367 cagttcattgctccagaaccagtgaaggtgtatccttgcaactggggaacacaaaagat 426
         RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
Db  1125 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRA 1066

Qy   427 t 427
           |
Db  1065 T 1065
```

NAME: Lynn E. Murry, Ph.D.
REGISTRATION NUMBER: 42,918
REFERENCE/DOCKET NUMBER: PA-0008 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 474 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNTUT02
CLONE: 614640
US-09-276-531-29

Query Match 2.5%; Score 35.4; DB 4; Length 474;
Best Local Similarity 48.3%; Pred. No. 0.11;
Matches 99; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

```
Qy   339 tgataagtcaagattcaccaatgttaaacagtcaattgctccagaaccagtgaaggtgt  398
          | ||   |  |  | ||   | |  | | |     |    |   ||  | |    ||
Db   359 TAACATTTGAAGAACCATACAACATAGACAACAAGCCTGTTCTTTCCCCTCCTTTTCTTT 300

Qy   399 atccttgcaactggggaacacaaaagatttattcattgacctcaaagttcttaac  458
          |   | |  | ||||| |  |   || | | |  |  || |  |  | |||
Db   299 TTAAGAAGAACTGGGCACCCCAAATACTATCTTTTAAATTCATCTCCAGTTCAGTTTCTCAG 240

Qy   459 aaatgaagcgtctgtggttctggagacgacagaaaagagatcagctcttcactgaca  518
          | | ||| | | |||  | |||| | | ||  || | |  ||| |  | || ||
Db   239 ATATCAAGAAGTATGTCTTGTCCTTTTGGGCCAAAGAAAAAAGTGAAGCCGGTTAACTTAGC 180

Qy   519 ttatgtctcaaataccagctaatt 543
          | | | |  || | || | | ||
Db   179 TTTAATGACTCAGTCCACACAGATT 155
```

RESULT 3
US-08-948-564-5/c
Sequence 5, Application US/08948564
Patent No. 6121512
GENERAL INFORMATION:
APPLICANT: Siminszky, Balazs
APPLICANT: Dewey, Ralph E.
APPLICANT: Corbin, Frederick T.
TITLE OF INVENTION: No. 6121512el Cytochrome P-450 Constructs and Methods of Producing Herbicide-Resistant Transgenic Plants
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Virginia C. Bennett
STREET: PO Box 37428
CITY: Raleigh
STATE: North Carolina
COUNTRY: USA
ZIP: 27627
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,564
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 5051-409
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-854-1400
TELEFAX: 919-854-1401

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
 LENGTH: 1644 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
 NAME/KEY: CDS
 LOCATION: 4..1542
US-08-948-564-5

```
Qy   399  atccttgcaactgggaacacaaaagatttattattcaattgacctcaagttcttaac  458
          ||| |||||||||||||   |||||||| |||| ||||||||||||  ||||||
Db   756  AACCTGTTCTCTGCGAACCTCCAAGGCGTTCTTTTCTTTCTTTCGCTTTGAGAAAAAGGGGCTTAG  697

Qy   459  aaatggaagcgtgtctgtggttctggagacgacagaaagaaatcagctcttca  511
          || |||||| | | ||   |  |||| || ||| |||||   ||   ||||||
Db   696  AATTGGAAGAGATAGTCATCAATTCTCGGGTCCAAAGTGATGAGAACACTCTTCA  644
```

RESULT 4
US-08-318-492-3/c
Sequence 3, Application US/08318492
Patent No. 5552312
GENERAL INFORMATION:
 APPLICANT: Lim, Bing
 APPLICANT: Adra, Chaker N.
 APPLICANT: Lelias, Jean-Michel
 TITLE OF INVENTION: RECOMBINANT HTM4 GENE, PROTEIN AND
  ASSAYS
 NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 STREET: Two Militia Drive
 CITY: Lexington
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02173
COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/318,492
 FILING DATE:
 CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
 NAME: Brook, David E.
 REGISTRATION NUMBER: 22,592
 REFERENCE/DOCKET NUMBER: BIH94-03
TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 861-6240
 TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
 LENGTH: 1661 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
 NAME/KEY: CDS
 LOCATION: 97..741
US-08-318-492-3

Query Match 2.4%; Score 33; DB 4; Length 1644;
Best Local Similarity 55.8%; Pred. No. 1.4;
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

---

RESULT 5
US-08-707-340-3/c
Sequence 3, Application US/08707340
Patent No. 5705615
GENERAL INFORMATION:
 APPLICANT: Lim, Bing
 APPLICANT: Adra, Chaker N.
 APPLICANT: Lelias, Jean-Michel
 TITLE OF INVENTION: RECOMBINANT HTM4 GENE, PROTEIN AND
  ASSAYS
 NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 STREET: Two Militia Drive
 CITY: Lexington
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02173
COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/707,340
 FILING DATE: 03-SEP-1996
 CLASSIFICATION: 536
PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/318,492
 FILING DATE: 06-OCT-1994
PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/675,648
 FILING DATE: 03-JUL-1996
ATTORNEY/AGENT INFORMATION:
 NAME: Brook, David E.
 REGISTRATION NUMBER: 22,592
 REFERENCE/DOCKET NUMBER: BIH94-03A2
TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 861-6240
 TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
 LENGTH: 1661 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
 NAME/KEY: CDS
 LOCATION: 97..738
US-08-707-340-3

Query Match 2.3%; Score 31.8; DB 1; Length 1661;
Best Local Similarity 56.1%; Pred. No. 3.4;
Matches 60; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy   454  ttaacaaatggaagcgtgtctgtggttctggagacgacagaaagaaatcagctcttcact  513
          ||| ||| ||||||   |  |  |  |||| || ||| |||||   ||   |||||| |||
Db  1352  TTTACATATCAGGGACATGTGTGAGTTCCAGCAGTTCCACTGAAATAAAATCCCTGAGCCGTT  1293
```

Qy   514 gtacattatgtctcaaaatacccagctaattgctttaaagaagagaaga 560
         |||||||||||  ||| ||||||| ||    ||  |||||| || || ||
Db  1292 AACCCTCCCATCTCAAATACAGTCGAAAATTTATTTAATTGATAATGA 1246

RESULT 6
US-08-994-578-3/c
Sequence 3, Application US/0899457B

Patent No. 5972688
GENERAL INFORMATION:
  APPLICANT: Lim, Bing
  APPLICANT: Adra, Chaker N.
  APPLICANT: Lellas, Jean-Michel
  TITLE OF INVENTION: Htm4 METHODS OF TREATMENT AND ASSAYS,
  TITLE OF INVENTION: AGONISTS AND ANTAGONISTS
  NUMBER OF SEQUENCES: 4
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
    STREET: Two Militia Drive
    CITY: Lexington
    STATE: Massachusetts
    COUNTRY: U.S.A.
    ZIP: 02173
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/994,578
    FILING DATE: December 19, 1997
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/707,340
    FILING DATE: 03-SEP-1996
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/675,648
    FILING DATE: 03-JUL-1996
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/318,492
    FILING DATE: 06-OCT-1994
  ATTORNEY/AGENT INFORMATION:
    NAME: Brook, David E.
    REGISTRATION NUMBER: 22,592
    REFERENCE/DOCKET NUMBER: BIH94-03A2Z
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (781) 861-6240
    TELEFAX: (781) 861-9540
  INFORMATION FOR SEQ ID NO: 3:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 1661 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    FEATURE:
      NAME/KEY: CDS
      LOCATION: 97..738

US-08-994-578-3

Query Match    2.3%;  Score 31.8;  DB 2;  Length 1661;
Best Local Similarity   56.1%;  Pred. No. 3.4;
Matches    60;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy   454 ttaacaaatggaagcgtgtctggttctgggacgacagaaaagaatcagctcttcact 513
         | |||| |   |||  | | | | ||||| | ||  || || || || |||| ||
Db  1352 TTTACATATCAGGGCCATGTGTGTGAGTTCCAGCAGTTCACTGAAATAAATCCTGAGCCTT 1293

Qy   514 gtacattatgtctcaaaatacccagctaattgctttaaagaagagaaga 560
         |||||||||||  ||| ||||||| ||    ||  |||||| || || ||
Db  1292 AACCCTCCCATCTCAAATACAGTCGAAAATTTATTTAATTGATAATGA 1246

RESULT 7
US-08-615-170-2
Sequence 2, Application US/08615170
Patent No. 5576776
GENERAL INFORMATION:
  APPLICANT: ORDAHL, Charles P.
  APPLICANT: AZAKIE, Anthony
  APPLICANT: MAR, Janet H.
  APPLICANT: FARRANCE, Iain K.G.
  APPLICANT: HALL, Deborah E.
  APPLICANT: STEWART, Alexandre F.R.
  APPLICANT: LARKIN, Sarah B.
  TITLE OF INVENTION: DTEF-1 ISOFORMS AND USES THEREOF
  NUMBER OF SEQUENCES: 32
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Townsend and Townsend Khourie and Crew
    STREET: Steuart Street Tower, One Market Plaza
    CITY: San Francisco
    STATE: California
    COUNTRY: US
    ZIP: 94105-1493
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/615,170
    FILING DATE:
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: PCT/US95/01526
    FILING DATE: 06-FEB-1995
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/191,493
    FILING DATE: 04-FEB-1994
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Heslin, James M.
    REGISTRATION NUMBER: 29,541
    REFERENCE/DOCKET NUMBER: 2307U-053120
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (415) 326-2400
    TELEFAX: (415) 326-2422
  INFORMATION FOR SEQ ID NO: 2:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 1627 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    MOLECULE TYPE: cDNA
    FEATURE:
      NAME/KEY: CDS
      LOCATION: 185..1480
    OTHER INFORMATION: /product= "TEF-1A"

US-08-615-170-2

Query Match    2.2%;  Score 31.4;  DB 1;  Length 1627;
Best Local Similarity   48.6%;  Pred. No. 4.6;
Matches    86;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy   392 aaggtgtatccttgcaactggggaacacaaaagaagatttattatttcattgacctcaagt 451
         || |||| || ||| || ||  | || | ||| ||| | || | ||| |||   || ||
Db  1119 ATGGTGTTCCAGTCAATATGAGAGCCCAGAAAATATGGTCATTACCGTTGTCCACTAAAG 1178

Qy   452 tcttaacaaatggaagcgtgtctgttctgggacgacagaaaagaatcagctcttcca 511
         | ||| |  ||||   | | |  | ||||| | ||  || | || || || |||| |
Db  1179 TGTGTTCCTTTGGAAAGCACAGTGGTGGAGAAGAAGTGGAGACAGAGTATGCACACTATGAAA 1238

Qy   512 ctgtacattatgtctcaaaatacccagctaattgctttaaagaagagacatatact 568

Db      1239 ATGGACACTATGCCTATCGCATTCATCGTTCTCCTCTCTGTGAATACACATGATAAACT 1295
             ||| |||| || |||     |||         ||| ||          |||    |||    ||| ||| |||

RESULT   8
US-08-615-170-4
; Sequence 4, Application US/08615170
; Patent No. 5776776
; GENERAL INFORMATION:
;   APPLICANT:  ORDAHL, Charles P.
;   APPLICANT:  AZAKIE, Anthony
;   APPLICANT:  MAR, Janet H.
;   APPLICANT:  FARRANCE, Iain K.G.
;   APPLICANT:  HALL, Deborah E.
;   APPLICANT:  STEWART, Alexandre F.R.
;   APPLICANT:  LARKIN, Sarah B.
;   TITLE OF INVENTION:  DTEF-1 ISOFORMS AND USES THEREOF
;   NUMBER OF SEQUENCES:  32
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Townsend and Townsend Khourie and Crew
;     STREET:  Steuart Street Tower, One Market Plaza
;     CITY:  San Francisco
;     STATE:  California
;     COUNTRY:  US
;     ZIP:  94105-1493
;   CURRENT APPLICATION DATA:
;     SOFTWARE:  Patentin Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/615,170
;     FILING DATE:
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/191,493
;     FILING DATE:  04-FEB-1994
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/US95/01526
;     FILING DATE:  06-FEB-1995
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  435
;     FILING DATE:
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Heslin, James M.
;     REGISTRATION NUMBER:  29,541
;     REFERENCE/DOCKET NUMBER:  2307U-053120
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (415) 326-2400
;     TELEFAX:  (415) 326-2422
;   INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1666 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  185..1519
;   FEATURE:
;     NAME/KEY:  misc_feature
;     LOCATION:  536..571
;     OTHER INFORMATION:  /product= "TEF-1B"
;   OTHER INFORMATION:  /product= "KTQVDKYDFSEK"
US-08-615-170-4

Query Match             2.2%;    Score 31.4;    DB 1;    Length 1666;
Best Local Similarity    48.6%;   Pred. No. 4.6;
Matches   86;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy       392 aaggtgtatcctgcaactggggaacacaaaagatttattatttcatttgacctcaagt 451
             |  ||||    ||        |||||||||        |||  |||||  ||  |  |||||  ||

---

Db      1158 ATGGTGTTTCCAGTCAATATGACAGCCCAGAAAATATGGTCATTACCTGTTCCACTAAG 1217
             |     ||||              |||    |||     ||       ||   |||     || |||

RESULT   9
US-08-781-891-209
; Sequence 209, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
;   APPLICANT:  Fu, Ying-Hui
;   APPLICANT:  Yu, Chang-En
;   APPLICANT:  Oshima, Junko
;   APPLICANT:  Mulligan, John T.
;   APPLICANT:  Schellenberg, Gerald D.
;   TITLE OF INVENTION:  GENE AND GENE PRODUCTS RELATED TO
;   TITLE OF INVENTION:  WERNER'S SYNDROME
;   NUMBER OF SEQUENCES:  209
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  SEED and BERRY LLP
;     STREET:  6300 Columbia Center, 701 Fifth Avenue
;     CITY:  Seattle
;     STATE:  Washington
;     COUNTRY:  USA
;     ZIP:  98104-7092
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/781,891
;     FILING DATE:  27-DEC-1996
;     CLASSIFICATION:  800
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  No. 6090620tenburg Ph.D., Carol
;     REGISTRATION NUMBER:  39,317
;     REFERENCE/DOCKET NUMBER:  240052.419
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (206) 622-4900
;     TELEFAX:  (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 209:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  51259 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
US-08-781-891-209

Query Match             2.2%;    Score 31.4;    DB 3;    Length 51259;
Best Local Similarity    50.3%;   Pred. No. 40;
Matches   77;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy       537 gctaattgcttttaaagagaaagagacatatactatgcatgggcccagaacaatggag 596
             |  |||         ||||||| |||     ||   ||  |||| ||  ||  ||||||||||

Db      8997 GATAATAAGGTATAGAAGAAAAGTGAAGAATTCCAACTCAAAAGGGCCAGTAAATATCTTCAA 9056

Qy       597 cacagttacccgggacctggtcactgaccccaggaaagaagtgggtctttccacacaa 656
             ||        |||||         ||  ||| |||  || |  ||||        ||| ||||

Db      9057 CAAAATTATAGAAAGAAAAACTTCCATAACCTAAAGAAAACGGATGTCCATGAACATACAAGA 9116

Qy       657 agctgtcaagccaacaagaataatgcccaagaa 689
             |  |||       |||| |||||||||   |||||

Db      9117 AACTTCCAGAACTCCAAATAGACTGGACAAGA 9149

RESULT 10
US-08-737-524B-26
Sequence 26, Application US/08737524B
Patent No. 5912414
GENERAL INFORMATION:
 APPLICANT: CARL SAVERIO FALCO
 APPLICANT: DOMINICK ANTHONY GUIDA, JR.
 APPLICANT: MARY ELIZABETH HARNETT LOCKE
 TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS, CHIMERIC
 TITLE OF INVENTION: GENES AND METHODS FOR INCREASING
 TITLE OF INVENTION: THE METHIONINE CONTENT OF THE SEEDS
 TITLE OF INVENTION: OF PLANTS
 NUMBER OF SEQUENCES: 27
 CORRESPONDENCE ADDRESS:
  ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
  STREET: 1007 MARKET STREET
  CITY: WILMINGTON
  STATE: DELAWARE
  COUNTRY: UNITED STATES OF AMERICA
  ZIP: 19898
 COMPUTER READABLE FORM:
  MEDIUM TYPE: DISKETTE, 3.50 INCH
  COMPUTER: IBM PC COMPATIBLE
  OPERATING SYSTEM: MICROSOFT WINDOWS 95
  SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (7.0)
 CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/737,524B
  FILING DATE:
  CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
  NAME: LYNNE M. CHRISTENBURY
  REGISTRATION NUMBER: 30,971
  REFERENCE/DOCKET NUMBER: BB-1059-A
 TELECOMMUNICATION INFORMATION:
  TELEPHONE: 302-992-5481
  TELEFAX: 302-773-0164
  TELEX: 835420
INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
  LENGTH: 3639 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
US-08-737-524B-26

Query Match    2.2%;    Score 31;    DB 2;    Length 3639;
Best Local Similarity    50.3%;    Pred. No. 10;
Matches    76;    Conservative    0;    Mismatches    75;    Indels    0;    Gaps    0;

Qy    426  tttattatttcatttgacctcaagtcttaacaaatggaaggacgcgtgtctgtggttctgga  485
               |||||| |||||||||||
Db   3200  TCTCGATATCACAAACCGCATCCAGTTCTCTGTTACATGAACTAGAATTCTGGTTCTGGA  3259

Qy    486  gacgacagaaagaaatcagctcttcactgtacattatgtctcaaatacccagctaattgc  545
               ||||||||||||||||||||||
Db   3260  AAGGAAGAAATAGGATATGTTCTGTGCACTGCAATATATATATCATTAATCCGAGC  3319

Qy    546  tctttaaagagaagagacatacttatggcatc  576
               ||
Db   3320  TTTATGTCACAGACTCACAGGCCAGGCTACC  3350

RESULT 11
US-08-147-812-4/c
Sequence 4, Application US/08147812
Patent No. 5766909
GENERAL INFORMATION:
 APPLICANT: Xie, Qiao-wen
 APPLICANT: Nathan, Carl F.
 APPLICANT: Munford, Richard A.
 APPLICANT: Calaycay, Jimmy Ramos

TITLE OF INVENTION: DNA Encoding Inducible Nitric Oxide Synthase
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
 ADDRESSEE: Merck & Co., Inc.
 STREET: 126 East Lincoln Avenue
 CITY: Rahway
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07065
COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy Disk
 COMPUTER: Macintosh Centris650
 OPERATING SYSTEM: Macintosh 7.0.1
 SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/147,812
 FILING DATE: No. 5766909 Available
 CLASSIFICATION: 435
PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/841,641
 FILING DATE: 02-FEB-1992
ATTORNEY/AGENT INFORMATION:
 NAME: Wallen, John W III
 REGISTRATION NUMBER: 35,403
 REFERENCE/DOCKET NUMBER: 18658IA
TELECOMMUNICATION INFORMATION:
 TELEPHONE: (908) 594-3905
 TELEFAX: (908) 594-4720
 TELEX: 138825
INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
  LENGTH: 4041 bases
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
US-08-147-812-4

Query Match    2.2%;    Score 30.8;    DB 1;    Length 4041;
Best Local Similarity    50.7%;    Pred. No. 13;
Matches    74;    Conservative    0;    Mismatches    72;    Indels    0;    Gaps    0;

Qy    613  ctggtcactgacctcaggaaaggagtcgggtctcttccaacacaaaagctgtcaagccaaca  672
               ||||| || || ||| ||||| | ||||| |  |||   ||||  | ||| ||  ||
Db   2166  CTGGTCGATGTCATGACCAAAGCGCAGAACTGAGGGTACATGCTGGAGCGCCAAGGCCAAA  2107

Qy    673  agaataatgcccaagaaagtggttaggtggttgattcggcgaaagggaagggcttccttgacaac  732
               || ||  ||||  |||| ||| ||||| |||| || |||
Db   2106  CACAGCATACCTGAAGGGTGGTTGAGTTCTCTAAGCATGAACAGAGAGATTTCTTCAGAGT  2047

Qy    733  attaccatctctaccacagcccacat  758
               ||  |||| ||  ||| ||||  |||
Db   2046  CTGCCCATTGCTGGGACAGTCTCCAT  2021

RESULT 12
US-09-123-708-1/c
Sequence 1, Application US/09123708
Patent No. 6146887
GENERAL INFORMATION:
 APPLICANT: SCHRADER, Juergen
 APPLICANT: GOEDECKE, Axel
 TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN GENE THERAPEUTIC
 TITLE OF INVENTION: TREATMENT OF VASCULAR DISORDERS
 FILE REFERENCE: 511169-2003
 CURRENT APPLICATION NUMBER: US/09/123,708
 CURRENT FILING DATE: 1998-07-28
 EARLIER APPLICATION NUMBER: 08/553,503
 EARLIER FILING DATE: 1996-03-01
 EARLIER APPLICATION NUMBER: P4411402.8
 EARLIER FILING DATE: 1994-03-31
 NUMBER OF SEQ ID NOS: 6
 SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1
; LENGTH: 4110
; TYPE: DNA
; ORGANISM: Cytomegalovirus
US-09-123-708-1

Query Match    2.2%;    Score 30.8;    DB 4;    Length 4110;
Best Local Similarity    50.7%;    Pred. No. 13;
Matches    74;    Conservative    0;    Mismatches    72;    Indels    0;    Gaps    0;

```
Qy   613 ctggtcactgacctcaggaaaggagtggggtcttccaacacaaaagctgtcaagccaaca 672
         |||||| |||||| ||||| ||||| | |||||
Db  2102 CTGGTCGATGTCATGAGCAAAAGCGCAGAACTGAGGGTACAATGCTGGAGCCAAGGCCAAA 2043
Qy   673 agaataatgcccaagaaggtggttaggttgattgcggaaaggaggggcttcccttgacaac 732
         | ||| | |||||||| ||| |
Db  2042 CACAGCATACCTGAAGGTGTGGTTGAGTTCTCTAAGCCATGAACACAGAGATTTCTTCAGAGT 1983
Qy   733 attaccatctctaccacagcccacat 758
         |||| | | ||  ||| |||
Db  1982 CTGCCCATTGCTGGGACAGTCTCCAT 1957
```

RESULT 13
US-09-123-624-1/c
; Sequence 1, Application US/09123624
; Patent No. 6149936
; GENERAL INFORMATION:
; APPLICANT: SCHRADER, Jurgen
; APPLICANT: GODECKE, Axel
; TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN THE GENE THERAPEUTIC
; TITLE OF INVENTION: TREATMENT OF VASCULAR DISORDERS
; FILE REFERENCE: 511169-2004
; CURRENT APPLICATION NUMBER: US/09/123,624
; CURRENT FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 08/553,503
; PRIOR FILING DATE: 1996-03-01
; PRIOR APPLICATION NUMBER: 4411402.8
; PRIOR FILING DATE: 1994-03-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 1
; LENGTH: 4110
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-123-624-1

Query Match    2.2%;    Score 30.8;    DB 4;    Length 4110;
Best Local Similarity    50.7%;    Pred. No. 13;
Matches    74;    Conservative    0;    Mismatches    72;    Indels    0;    Gaps    0;

```
Qy   613 ctggtcactgacctcaggaaaggagtggggtcttccaacacaaaagctgtcaagccaaca 672
Db  2102 CTGGTCGATGTCATGAGCAAAAGCGCAGAACTGAGGGTACAATGCTGGAGCCAAGGCCAAA 2043
Qy   673 agaataatgcccaagaaggtggttaggttgattgcggaaaggaggggcttcccttgacaac 732
Db  2042 CACAGCATACCTGAAGGTGTGGTTGAGTTCTCTAAGCCATGAACACAGAGATTTCTTCAGAGT 1983
Qy   733 attaccatctctaccacagcccacat 758
Db  1982 CTGCCCATTGCTGGGACAGTCTCCAT 1957
```

RESULT 14
US-08-147-812-6/c
; Sequence 6, Application US/08147812
; Patent No. 5766909
; GENERAL INFORMATION:
; APPLICANT: Xie, Qiao-wen
; APPLICANT: Nathan, Carl F.
; APPLICANT: Mumford, Richard A.
; APPLICANT: Calaycay, Jimmy Ramos
; TITLE OF INVENTION: DNA Encoding Inducible Nitric Oxide Synthase
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 East Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh Centris650
; OPERATING SYSTEM: Macintosh 7.0.1
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,812
; FILING DATE: No. 5766909 Available
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/841,641
; FILING DATE: 02-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wallen, John W III
; REGISTRATION NUMBER: 35,403
; REFERENCE/DOCKET NUMBER: 18658IA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3905
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4165 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-147-812-6

Query Match    2.2%;    Score 30.8;    DB 1;    Length 4165;
Best Local Similarity    50.7%;    Pred. No. 13;
Matches    74;    Conservative    0;    Mismatches    72;    Indels    0;    Gaps    0;

```
Qy   613 ctggtcactgacctcaggaaaggagtggggtcttccaacacaaaagctgtcaagccaaca 672
Db  2166 CTGGTCGATGTCATGAGCAAAAGCGCAGAACTGAGGGTACAATGCTGGAGCCAAGGCCAAA 2107
Qy   673 agaataatgcccaagaaggtggttaggttgattgcggaaaggaggggcttcccttgacaac 732
Db  2106 CACAGCATACCTGAAGGTGTGGTTGAGTTCTCTAAGCCATGAACACAGAGATTTCTTCAGAGT 2047
Qy   733 attaccatctctaccacagcccacat 758
Db  2046 CTGCCCATTGCTGGGACAGTCTCCAT 2021
```

RESULT 15
US-08-701-582D-1
; Sequence 1, Application US/08701582D
; Patent No. 6017755
; GENERAL INFORMATION:
; APPLICANT: WRANA, Jeffrey
; APPLICANT: ATTISANO, Liliana
; APPLICANT: SCHERER, Stephen W.
; TITLE OF INVENTION: MADR2 TUMOUR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.

```
ZIP:  22313-1404
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER:  IBM PC compatible
  OPERATING SYSTEM:  PC-DOS/MS-DOS
  SOFTWARE:  PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/701,582D
  FILING DATE:  22-AUG-1996
  CLASSIFICATION:  435
ATTORNEY/AGENT INFORMATION:
  NAME:  REA, Teresa Stanek
  REGISTRATION NUMBER: 30,427
  REFERENCE/DOCKET NUMBER:  024916-004
TELECOMMUNICATION INFORMATION:
  TELEPHONE:  (703) 836-6620
  TELEFAX:  (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
  LENGTH:  1605 base pairs
  TYPE:  nucleic acid
  STRANDEDNESS:  single
  TOPOLOGY:  linear
  MOLECULE TYPE:  cDNA
US-08-701-582D-1
```

```
Query Match      2.2%;   Score 30.6;  DB 3;   Length 1605;
Best Local Similarity   49.1%;   Pred. No. 8.3;
Matches  81;   Conservative  0;   Mismatches  84;  Indels  0;  Gaps  0;

Qy   1024 atgaataaacatgactggtatgaagaaatatccaactaacctagcctcttttgtttttaaat 1083
            |  |||  |   ||  |||  || | ||  |   || ||| | ||| |||| | ||| |||
Db    229 ATCACCACTCAAAACTGTAATACTAAATGTGTTACCATACCAAGCACTTGCTCTGAAATT 288

Qy   1084 ggctttatgtattcttttaattgggctgtatgacttaaaagaaactgcagggaaaactc 1143
            ||  |  | ||   | ||  | |||    | | || |||||   | ||  || | |
Db    289 TGGGGACTGAGTACCACCAAATACGATAGATCAGTGGGATACAACAGGCCTTTACAGCTTC 348

Qy   1144 gggaaagaagcgaggtccttgtatgagcgtggcatggaatccctt 1188
            || ||| || ||||  || |  ||  |||| | ||  |||||||
Db    349 TCTGAACAAACCAGGTCTCTTGATGGTCGTCCTCCAGGTATCCCAT 393
```

OM nucleic - nucleic search, using sw model

                    GenCore version 4.5
                    Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:             May 25, 2001, 01:57:43 ; Search time 1044.13 seconds
                                            (without alignments)
                                            11747.083 Million cell updates/sec

Title:              US-09-403-269-12_COPY_1_1404
Perfect score:      1404
Sequence:           1 tccaagctgaattctcatag..........gcagccgggcaaagcacaac 1404

Scoring table:      IDENTITY_NUC
                    Gapop 10.0 , Gapext 1.0

Searched:           9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters:      19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:    Minimum Match 0%
                    Maximum Match 100%
                    listing first 45 summaries

Database :

EST:*

| | | |
|---|---|---|
| 1: gb_est1:* | 44: em_esthum10:* | 88: em_estro7:* |
| 2: gb_est2:* | 45: em_esthum11:* | 89: em_estro8:* |
| 3: gb_est3:* | 46: em_esthum12:* | 90: em_estro9:* |
| 4: gb_est4:* | 47: em_esthum13:* | 91: em_estro10:* |
| 5: gb_est5:* | 48: em_esthum14:* | 92: em_estro11:* |
| 6: gb_est6:* | 49: em_esthum15:* | 93: em_estro12:* |
| 7: gb_est7:* | 50: em_esthum16:* | 94: em_estro13:* |
| 8: gb_est8:* | 51: em_esthum17:* | 95: em_estro14:* |
| 9: gb_est9:* | 52: em_esthum18:* | 96: em_estro15:* |
| 10: gb_est10:* | 53: em_esthum19:* | 97: em_estro16:* |
| 11: gb_est11:* | 54: em_esthum20:* | 98: em_estro17:* |
| 12: gb_est12:* | 55: em_esthum21:* | 99: em_estro18:* |
| 13: gb_est13:* | 56: em_esthum22:* | 100: em_estro19:* |
| 14: gb_est14:* | 57: em_esthum23:* | 101: em_estro20:* |
| 15: gb_est15:* | 58: em_esthum24:* | 102: gb_est25:* |
| 16: gb_est16:* | 59: em_esthum25:* | 103: gb_est26:* |
| 17: gb_est17:* | 60: em_esthum26:* | 104: gb_est27:* |
| 18: gb_est18:* | 61: em_esthum27:* | 105: gb_est28:* |
| 19: gb_est19:* | 62: em_esthum28:* | 106: gb_est29:* |
| 20: gb_est20:* | 63: em_estin1:* | 107: gb_est30:* |
| 21: gb_est21:* | 64: em_estin2:* | 108: gb_est31:* |
| 22: gb_est22:* | 65: em_estin3:* | 109: gb_est32:* |
| 23: gb_est23:* | 66: em_estin4:* | 110: gb_est41:* |
| 24: gb_est24:* | 67: em_estin5:* | 111: gb_est42:* |
| 25: gb_est33:* | 68: em_estom1:* | 112: gb_est43:* |
| 26: gb_est34:* | 69: em_estom2:* | 113: gb_est44:* |
| 27: gb_est35:* | 70: em_estov1:* | 114: gb_est45:* |
| 28: gb_est36:* | 71: em_estov2:* | 115: gb_est46:* |
| 29: gb_est37:* | 72: em_estpl1:* | 116: gb_est47:* |
| 30: gb_est38:* | 73: em_estpl11:* | |
| 31: gb_est39:* | 74: em_estpl12:* | |
| 32: gb_est40:* | 75: em_estpl13:* | |
| 33: em_estba:* | 76: em_estpl14:* | |
| 34: em_estfun:* | 77: em_estpl15:* | |
| 35: em_esthum1:* | 78: em_estpl16:* | |
| 36: em_esthum2:* | 79: em_estpl17:* | |
| 37: em_esthum3:* | 80: em_estpl18:* | |
| 38: em_esthum4:* | 81: em_estpl19:* | |
| 39: em_esthum5:* | 82: em_estpl10:* | |
| 40: em_esthum6:* | 83: em_estro1:* | |
| 41: em_esthum7:* | 84: em_estro2:* | |
| 42: em_esthum8:* | 85: em_estro3:* | |
| 43: em_esthum9:* | 86: em_estro4:* | |
| | 87: em_estro5:* | |
| | 88: em_estro6:* | |

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*

190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| | Result No. | Score | Query Match % | Length | DB ID | Description |
|---|---|---|---|---|---|---|
| | 1 | 599.6 | 42.7 | 698 | 103 | AI860635 | AI860635 w104a09.x |
| | 2 | 586.2 | 41.8 | 776 | 107 | AU139373 | AU139373 AU139373 |
| c | 3 | 521 | 37.1 | 653 | 114 | AW383385 | AW383385 PM3-HT034 |
| c | 4 | 450.4 | 32.1 | 559 | 21 | AI521253 | AI521253 to66h09.x |
| | 5 | 401.8 | 28.6 | 477 | 114 | AW383386 | AW383386 PM3-HT034 |
| c | 6 | 364.2 | 25.9 | 426 | 7 | AA435342 | AA435342 vd49g09.s |
| | 7 | 346.6 | 24.7 | 455 | 12 | AA794925 | AA794925 vr48g04.s |
| | 8 | 331.6 | 23.6 | 866 | 230 | CNS01WIY | AL170467 Tetraodon |
| c | 9 | 304.2 | 21.7 | 358 | 20 | AI467374 | AI467374 vd49g09.x |
| | 10 | 303 | 21.6 | 434 | 163 | BE170776 | BE170776 QV4-HT053 |
| c | 11 | 285.4 | 20.3 | 566 | 143 | BF080558 | BF080558 231598 MA |
| | 12 | 285.4 | 20.3 | 590 | 104 | AI981391 | AI981391 pat.PK005 |
| | 13 | 195 | 13.9 | 381 | 140 | BE861166 | BE861166 UI-M-AM0- |
| | 14 | 164.6 | 11.7 | 251 | 160 | BB579985 | BB579985 BB579985 |
| | 15 | 143.6 | 10.2 | 684 | 31 | AV672171 | AV672171 AV672171 |
| c | 16 | 139.4 | 9.9 | 203 | 117 | AW605325 | AW605325 QV3-DT004 |
| c | 17 | 129.2 | 9.2 | 663 | 31 | AV676110 | AV676110 AV676110 |
| c | 18 | 104.8 | 7.5 | 802 | 230 | CNS02E30 | AL193221 Tetraodon |

RESULT 1

LOCUS       AI860635/c        698 bp    mRNA            EST        07-MAR-2000
DEFINITION  wl04a09.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2423896 3',
            mRNA sequence.
            similar to TR:O18756 O18756 C5-GLUCURONYL EPIMERASE ;, mRNA
            sequence.
ACCESSION   AI860635
VERSION     AI860635.1       GI:5514251
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 698)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq Primer: -40UP from Gibco
            High quality sequence stop: 420.
            Insert Length: 950    Std Error: 0.00
            Seq Primer: -40UP from Gibco
FEATURES             Location/Qualifiers
     source          1..698
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2423896"
                     /clone_lib="NCI_CGAP_Ut1"
                     /tissue_type="well-differentiated endometrial
                     adenocarcinoma, 7 pooled tumors"
                     /lab_host="DH10B"

| c | 19  | 102.8 | 7.3 | 31  | CNS02GE  | AV677068 |
| c | 20  | 88.4  | 6.3 | 230 | CNS02GE  | AV677068 |
| c | 21  | 79.4  | 5.7 | 172 | BG019834 | AL196007 Tetraodon |
| c | 22  | 78    | 5.6 | 568 | BG019834 | dc68h03.x |
| c | 23  |       | 5.6 | 1078| CNS039E3 | Tetraodon |
| c | 24  | 62    | 4.4 | 951 | CNS0296I | Tetraodon |
| c | 25  | 43    | 3.1 | 609 | BE325397 | AI658400 AEMTBK54 |
| c | 26  | 40.2  | 2.9 | 473 | BE325397 | NF087E08S |
| c | 27  | 40.2  | 2.9 | 1201| CNS0165X | AL106335 Drosophil |
| c | 28  | 38.8  | 2.8 | 271 | BB149156 | BB149156 |
| c | 29  | 38.6  | 2.8 | 1101| CNS016GZ | AL106733 Drosophil |
| c | 30  | 38.6  | 2.7 | 196 | CNS040GX | AZ734735 Drosophil |
| c | 31  | 38.4  | 2.7 | 789 | CNS03VTA | AZ734735 RPCI-24-1 |
| c | 32  | 38.2  | 2.7 | 907 | BF529053 | AL262855 Tetraodon |
| c | 33  | 38.2  | 2.7 | 562 | BE942638 | BF529053 |
| c | 34  | 38.2  | 2.7 | 1101| CNS003FR | BE942638 EST422217 |
| c | 35  | 38    | 2.7 | 501 | CNS017YY | AL064653 Drosophil |
| c | 36  | 38    | 2.7 | 507 | B95711   | AL108676 Drosophil |
| c | 37  | 38    | 2.7 | 873 | CNS033MS | B95711 CIT-HSP-216 |
| c | 38  | 37.8  | 2.7 | 392 | BF935500 | AI944742 bs04h03.x |
| c | 39  | 37.6  | 2.7 | 975 | CNS04RQ0 | AL226333 Tetraodon |
| c | 40  | 37.4  | 2.7 | 238 | AV246889 | BF935500 CM1-NT027 |
| c | 41  | 37.4  | 2.7 | 246 | BB064841 | AL304209 Tetraodon |
| c | 42  | 37.4  | 2.7 | 124 | BB064841 | AV246889 AV246889 |
| c | 43  | 37.2  | 2.6 | 1004| CNS040GX | AL268890 Tetraodon |
| c | 44  | 36.8  | 2.6 | 232 | AQ747277 | AQ747277 HS_5537_A |
| c | 45  | 36.8  | 2.6 | 522 | AV239910 | AV239910 se68e11.y |
| c |     | 36.8  | 2.6 | 112 | AW186496 | AW186496 |

ALIGNMENTS

BASE COUNT      184 a    152 c    140 g    219 t    3 others
ORIGIN          /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.75 kb. Life Technologies catalog #:
                11538-014"

Query Match    42.7%;  Score 599.6;  DB 103;  Length 698;
Best Local Similarity   91.1%;  Pred. No. 1.4e-166;
Matches  635;  Conservative  0;  Mismatches  62;  Indels  0;  Gaps  0;

Qy    359 atgttaaacagttcattgcttccagaaaaccagtgaaggtgtatcttgcaactgggaaca 418
          ||||| ||||| |||||||||||||||||| || |||||||||||||||||| |||||||
Db    697 ArgTCAAACAGTTAAAttgCACCAAAAAAACAACTGAGGGtGATtCttTTGCAACtGGGA...AACAC 638

Qy    419 caaaagagtttattattcattgacctcaagtttcttaacaaatggaagcvgtgtctgtgg 478
          ||       ||||||||||||||||| || |||||||||||||||||| ||||||||||
Db    637 AAANGATTTTTATTATTCATTgacctcAAGTtCTTGACAAAtGNAACtGTGTCCGTGG 578

Qy    479 ttcctggagacgacagaaaagaatcagcttcttcactgtaccatttgtctcaaatacccagc 538
          |||| || |||||| ||| |||||| ||||||||||||||| ||||||| ||||||||||||
Db    577 TTCTAGAGACGACCACAAAAAGAATCAGCTCTTCACTATACATTTATGTCTCAAAAtACCAGC 518

Qy    539 taattgtcttttaaagaaagagaacatatactatgggatcgggcccagaacaatcatgggaca 598
          ||||||||||||||||||||||||||||||||||||||||| || |||||||||||| ||||
Db    517 TAATTGCTTTTAAAGAAAAAGATAtATAtACTAtGGCATTGGGCCCCAGAACttCAtGGGAGCA 458

Qy    599 cagcttaccgggaccttggtgaggaacccaggatgaagaaaggvcggctggggattatggvtgacccgta 658
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    457 CAGTTACCAGGGACCTTGGtCGtCACTGACCTCAGGAAAGGAGTGGGGTTCTTTCAAACACAACAAAG 398

Qy    659 ctgtcaagccaacaagaataatgcccaagaaagttgggttgttgatttgatttgcgaaagggaagg 718
          ||| |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    397 CTGTCAAGCCAACCAACAAAATAATGCCCAAGAAAGGTGGTTAGGTGATtCTTGCAACTGGAAAACAC 338

Qy    719 gcttccttgacacacattccacattctcctaccacagcccacacatggctgcctctctctcgctgcca 778
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    337 GATTCCTCCGACAACATTACCATCTCTACCACACAGCCCACATGGCTGCATTTTTTGCTGCTA 278

Qy    779 gtggactggctgtgaggaaccaggatgaagaaaggcggctggccgattatggtgaccgta 838
          ||| |||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    277 GTGAPTGGCTGAGTAAGGAACCAGGATGAGAAAAGGTGGCTGCCAAtTTAtGGTGACCGGTA 218

Qy    839 agttagggagaaggcttcaagtcttcagggttggtactccgcccatggccaagggc 898
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    217 AGTTAGGGAGGAAGGTTCAAGTCTTCTTTAGAGCCAGGATGGAATTCTGCAGAGGGC 158

Qy    899 aagccaattctacaattagtcagggcctatcctcttaacaaaagacaatatattcctcaatt 958
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 AAGCCAtTTCTACAtTAGTCAGGGCCTAtCTCGTTAACAAAAGACCAtTAtATtCCTCAAT 98

Qy    959 cagcttaaggcaacagccctacaagtttcctgtcaagagacagcatggaagtcaaggctg 1018
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     97 CAGCTTTAAGGCGGCAACAGCGCCCTTATAAGTTTCTATCTGACCAGCACATGGAGTCTAAAGCTG 38

Qy   1019 tgtttatgaataaacatgactggtatgaaaatacc 1055
          ||||||||||||||||||||||||||||||||||||
Db     37 TGTTTATGAATAAAACATGACTGGTATGAAGAATATCC 1

RESULT 2

LOCUS       AU139373          776 bp    mRNA            EST        25-OCT-2000
DEFINITION  AU139373 PLACE1 Homo sapiens cDNA clone PLACE1010481 5', mRNA
            sequence.
ACCESSION   AU139373
VERSION     AU139373.1       GI:11000894
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1  (bases 1 to 776)
AUTHORS     Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
TITLE       HRI human cDNA project
JOURNAL     Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..776
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1010481"
                     /clone_lib="PLACE1"
                     /tissue_type="placenta"
                     /note="Vector: pME18SFL3"
BASE COUNT      248 a     151 c     177 g     197 t      3 others
ORIGIN

Query Match              41.8%;   Score 586.2;  DB 107;  Length 776;
Best Local Similarity    92.7%;   Pred. No. 1.4e-162;
Matches  659;  Conservative   0;  Mismatches  45;  Indels   7;  Gaps  4;

Qy     13 tctcatagctcattccaaagtctttatggtcacagagagagccccttatccacctgatgtgtgttt 72
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     53 TCTCATAGCTCATTCCAAAGTCTATGCACAGAGAGCCCCTTATCCACCTGATGGTGTGTTT 112

Qy     73 atgtcctttgaaggtctacaatgtgaagtccgagacagagccgaagtgtcataagtgggtt 132
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113 ATGTCTTTTGAAGGCTACAAATGTGGAAGTCCGAGACAGAGTCGAAGTGCATAAGTGGGTT 172

Qy    133 gaaggtgtacctttatctcacacagtcaaggtcatttcctcaccaatccagatt 192
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    173 GAAGGTGTGCCATTAACTACACAATGGGACCTCAAGGTGTATTTCTATCCAGATT 232

Qy    193 gcacagtatgggttaagtcactacacagcaagagaatctaactgaaaaacccctcatatagag 252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    233 GCACAGTATGGATTAAGTCAATTACACAGAATCTAACTGAGAAACCTCTCACATAGAG 292

Qy    253 gtatatagaaacagcagaagaacgagggacaaaaacaacagcccaatgactggtgtgtgccc 312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    293 GTATATGAAACAGCAGAGAGCAGAGACAGACAAAAA---CAAGCCTAATGACTGGACTGTGCCA 349

Qy    313 aagggctgccttttatgggctagttgtggctgataagtcaagattcaccaatgttaaacagttc 372
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    350 AAGGGCTGCTTTATGGGGCGAATGTGGCGTCATAAGTCTAGATTCACCAATGTCAAACAGTTT 409

Qy    373 atgctccagaaaccagtgaaggtgtatcctttgcacatgggggaacacaaaagatttatt 432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    410 ATTGCACCAGAAACCAGTGAAGGTGTATCCTTGCAACTGGGAAACACAAAAGAGATTTATT 469

Qy    433 atttcatttgacctcaagttcttaacacaatggaagcgtgtctgttgttct-gggagacgac 491
          ||||||||||||||||||||||||||||||||||||||||||||||||||  |||||||||
Db    470 ATTTCATTTGACCTCAAGTTCTTGACAAAATGGAAGTGTGTCTGTTTCTCGGTGAAGACCAC 529

Qy    492 agaaaagaatcagctcctccacgtacatcattatgtctcaaataccacagctaattgcttttaa 551
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    530 AGAAAAGAATCAGCTCCTCACTACATCATTATGTCTCAATATCGTCAGCTAATTGCTTTTAA 589

Qy    552 agaaagagacacatatactat-ggcatcgggcccagaacatcatcatggaggacagttacccggg 610
          |||||||||||||||||| || ||||||||||||||||||||||||||||||||||||||
Db    590 AGAAAGAGATAATAATACTATGGGCATGGCAGAAACTTCATGGAGGACAGTTACCCAGGG 649
```

---

```
RESULT   3

LOCUS       AW383385     653 bp     mRNA      EST    04-FEB-2000
DEFINITION  PM3-HT0347-091199-001-b06 HT0347 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW383385
VERSION     AW383385.1  GI:6888044
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 653)
AUTHORS     HCGP http://www.ludwig.org.br/ORESTES.
TITLE       The FAPESP/LICR Human Cancer Genome Project
JOURNAL     Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Ludwig Institute for Cancer Genetics
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-27049922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?til=PM3&t2=PM3-HT0347-
            091199-001-b06&t3=1999-11-09&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 38
            High quality sequence stop: 567.
FEATURES             Location/Qualifiers
     source          1..653
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HT0347"
                     /clone_lib="HT0347"
                     /dev_stage="Adult"
                     /note="Organ: head,neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      215 a     131 c     143 g     164 t
ORIGIN

Query Match              37.1%;   Score 521;  DB 114;  Length 653;
Best Local Similarity    92.1%;   Pred. No. 2.9e-143;
Matches  572;  Conservative   0;  Mismatches  45;  Indels   4;  Gaps  2;

Qy    183 aatccagattgcacagtatgggttaagtcactacacagcaagaaatctaactgaaaaacccccc 242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8 AATCCAGATTGCACAGTATGCACAGTATGGATTAAGTCATTACACAGCAAGAATCTAACTGAGAAACCTCC 67

Qy    243 tcatatagaggtatatgaaacagcagaagaacgagggacaaaaacaacagcccaatgactg 302
          ||||| | ||||||||||||||||||||||||||||||||||||||||||||||||
Db     68 TCACATAGAGGTATATGAAACAGCAGAGAGCAGAGACAGACAAAAA---CAAGCCTAATGACTG 124

Qy    303 gactggtgccaagggctgctttatggtggctgataagtcaagattcaccaatgt 362
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 GACTGTGCCAAAGGGCTGCTTTATGGGGCGAATGTGGCGTCATAAGTCTAGATTCACCAATGT 184

Qy    363 taaacagttcattgctccagaaaccagtgaaggtgtatcctttgcaactgggggaacacaaa 422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 ...
```

---

```
Qy    611 acctggtcaactgacctccaggagaaggaggagtgggtctcttccaacaacaaaagtgtcaaagccaa 670
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    650 ACCTGGTCAGTCGACCTCAGGAAGAGGACGGAGTGGGTGTCTTTCAAACACAAAAGCTGTCAAAGCCAA 709

Qy    671 caagaataatgcccaag--aaggtggttaagttgattgccgaaaggaaggg 719
          ||||| |||||   |||  |||||||||||||||  ||||||| ||  |||
Db    710 CCAANATAATGCCCAAGNAAAGGTGGTNAGGTTGGTTAGGTTGATTTGCAAAAAGGTAAGG 760
```

RESULT 4

LOCUS       AI521253/c

LOCUS       AI521253    559 bp    mRNA    EST    12-MAY-1999
DEFINITION  to66h09.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2183297 3',
            mRNA sequence.
            similar to TR:O18756 O18756 C5-GLUCURONYL EPIMERASE ;.  mRNA
            sequence.
ACCESSION   AI521253
VERSION     AI521253.1  GI:4435388
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 559)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: Robert_Strausberg@nih.gov
            Tel: (301) 496-1550
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1602    Std Error: 0.00
            Seq Primer: -40UP from Gibco
            High quality sequence stop: 376
            POLYA=NO.
FEATURES             Location/Qualifiers
     source          1..559
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2183297"
                     /clone_lib="NCI_CGAP_Gas4"
                     /tissue_type="poorly differentiated adenocarcinoma with
                     signet ring cell features"

```
Db  185  CAAACAGTTTATTTGCACCAGAAACCAGTGAAGGTGTATCCTTGCAACTGGGAAACACAAA
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  423  agatttattattcattcaattgaacctcaaagttctgtgttct  482
Db  245  AGATTTATTATTTCATTTGACCTCAAGTTCTTGACAAGTTCTTGACAAATGGGAAGTGTCCGTGGTTCT  304
Qy  483  ggagacgacaagaaagatcagctcttcacttgtacattatgtctcaaataccagctaat  542
Db  305  AGAGACACAGAAAGAAGAATCAGCTCTTCACTATACATTATGTCTCAAATGCCTCAGCTAAT  364
Qy  543  tgtctttaaagaagacaatacactatggtcatccgggcccagaacatcatggacacagt  602
Db  365  TGCTTTTAAAGAAGAAACAGAGATATATACTATGGCANTTGGGCCCAGAAACTTCATGGACACAGT  424
Qy  603  tacccgggacctggtcactgacctccaggaaaggagtgtctttccaacacaaaagctgt  662
Db  425  TACCAGGGACCTGGTCACTGACCTCCAGGAAAGGAGTGGGTCTTCAAACACACAAAAGCTGT  484
Qy  663  caagccaacaagaataatgcccaagaagaggtggttaggttgcgaaaggaagggctt  722
Db  485  CAAGCCAACCAAAATAATAACCCAAGAAGGTGGTTAGGTTGGATTGCCAAAAGGTAAGGGAATT  544
Qy  723  ccttgacaacattaccatctctacacaagcccacatgggctgc-ctcttcgctgccagtg  781
Db  545  CCTCGACAACATTACCATCTCTACACACAGCCCACATGGGTTGCANTTTTTTGCTCGCTAGTG  604
Qy  782  actggctggtgaggaaccagg  802
Db  605  AATTGGCTAGTAAGGAACCACG  625
```

RESULT 5

LOCUS       AW383386/c

LOCUS       AW383386    477 bp    mRNA    EST    04-FEB-2000
DEFINITION  PM3-HT0347-091199-001-b10 HT0347 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW383386
VERSION     AW383386.1  GI:6888045
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 477)
AUTHORS     HCGP http://www.ludwig.org.br/ORESTES.
TITLE       The FAPESP/LICR Human Cancer Genome Project
JOURNAL     Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
                     /lab_host="DH10B"
                     /note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT;
                     Average insert size 1.69 kb. Life Technologies catalog #:
                     11549-011."

BASE COUNT    145 a    130 c    113 g    169 t    2 others
ORIGIN

Query Match
Best Local Similarity    90.9%;
Matches   511;  Conservative    0;  Score 450.4;  Pred. No. 2.4e-122;  DB 21;  Length 559;
Mismatches   48;  Indels    3;  Gaps   3;

```
Qy  490  acagaaaagaataacagctcttcactgtacattatgtctcaaatacccagctaattgtttt  549
         ||||||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||
Db  559  ACAGAAAAGAATCAACCTTCTTCACTATCACTATACATTATGTCTCAAATGCCTCAGCTAATTGCTTTT  500
Qy  550  aaagaagagacatatactatggcatcgggcccagaacaatcatggacacagttacccgg  609
Db  499  -AAGAAGAGAGATATATACTATGGCATTGGGCCCAGAACTTCATGGACACAGTTACCCGG  441
Qy  610  gacctggtcactgacctccaggaagaggagtggttaggttggttgtcttccacacaaaagcca  669
Db  440  GACCTGGTCACTG-CCTCAGGAAAGGAGTGGGTCTTCAAACACACAAAAGCTGTCAAGCCA  382
Qy  670  acaagaataatgcccaagaagaggtgggttaggttgattgcgaaaggaagggctccttgac  729
Db  381  ACCAAAATAATAACCCAAGAAGGTGGTTAGGTTGCAAAAGGTAAGGGAATTCCTCGAC  322
Qy  730  aacattaccatctctacacagcccacatggctgcttcttcgctgccagtgactggctg  789
Db  321  -ACATTACCATCTCTACACACAGCCCACATGGGTTGCANTTTTTTGCTCGCTAGTGAATTGGCTA  263
Qy  790  gtgaggaaccaggatgagaaaggcggctggccgattatggtgaccgtaagttaggggaa  849
Db  262  GTAAGGAACCACGAAGATGAGAAAAGTGGCTGGCCAATTATGGTGACCGTAAGTTAGGGGAA  203
Qy  850  ggcttcaagtcttagagccaagggtggtactccgccatggcccaagggcaagccaattct  909
Db  202  GGGTCTCAAGTCTTTAGAGCCAAGGGTATTCTCCGCCATGGCCCAAGGGCAAGGCCATTTCT  143
Qy  910  acattagtcaggcctatctctaacaaagaccatatttattcctcaattcagcttaagg  969
Db  142  ACATTAGTCAGGGCCTATCTGTTAACAAAAGACCATATATTCCTCAATTCAGCTTTAAGG  83
Qy  970  gcaacagccccttacaaagtttctgtcaggcagcatggagtcaaggcctgttgtttatgaat  1029
Db   82  GCAACAGCCCCTTATAAGTTCTATCTGAGCAGCAGCATGGAGTTAAAGCTGTGTTTATGAAT  23
Qy 1030  aaacatgactggtatgaagaat  1051
Db   22  AAACATGACTGGTATGAAGAAT  1
```

## Left column

```
                  Brazil
                  Tel: +55-11-2704922
                  Fax: +55-11-2707001
                  Email: asimpson@ludwig.org.br
                  This sequence was derived from the FAPESP/LICR Human Cancer Genome
                  Project. This entry can be seen in the following URL
                  (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&t2=PM3-HT0347-
                  091199-001-b16&t3=1999-11-09&t4=1)
                  Seq primer: puc 18 forward
                  High quality sequence stop: 338.

FEATURES          Location/Qualifiers
source            1..477
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone_lib="HT0347"
                  /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                  Site_2: SmaI; A mini-library was made by cloning products
                  derived from ORESTES PCR (U.S. Letters Patent application
                  No. 196,716 - Ludwig Institute for Cancer Research)
                  profiles into the pUC 18 vector. Reverse transcription of
                  tissue mRNA and cDNA amplification were performed under
                  low stringency conditions."
                  /dev_stage="Adult"

BASE COUNT        122 a    115 c    102 g    138 t
ORIGIN

Query Match  92.9%;   Score 401.8;   DB 114;   Length 477;
Best Local Similarity  28.6%;   Pred No. 6e-108;
Matches  443;   Conservative  0;   Mismatches  32;   Indels  2;   Gaps  2;

Qy   517 cattatgtctcaaatacccagctaattgctttttaaagaaagaaagaacatatactatgggcatc 576
         ||||||||||||| |||||| | |||||||||||| ||||||||||| || ||||||||| |||||
Db   477 CATTATGTCTCAAAATGCTCAAGCTAATTGCTTTTTGAAGAAAGAGATATATACTATGGGCATT 418

Qy   577 gggcccagaacatcatggagacacagttacccggacctcggtcacctgacctcaggaaagg 636
         ||||||||||| ||||||| ||||||||||||| ||||| |||||||||| ||||||||||
Db   417 GGGCCCAGAACTTCATGGAGCACACAGTTACCACGGACCTCGGTCACTCAGCGAAAGGA 358

Qy   637 gtgggtcttttccaacacaacaaaagctgtcaagccaacaacaagaagaaggtgtgtt 696
         |||||||||| |||| ||||  ||||||| ||||| ||||||||||||||||||
Db   357 GTGGGTCTTTTCGTCACGAAAGCTGTCAACCAACAACAAAATAAYGCCCAAGAAGGTGGTT 298

Qy   697 aggttgaattgcgaaaggggaagggggcttccttgacaacaattaccatctctaccagcccac .756
         |||||||| || ||||||||||||| ||||| ||||||||||||||||||||||| |||||||
Db   297 AGGTTGATTGCAAAAGGTAAGGGATTCCTCGACGACACATTACCATCTCTACCACAGCCCAC 238

Qy   757 atggctgc-ctccttcgctcgcagtcgactcggtcggtggaaccaggagtgagaagaaggcgg 815
         |||||||| |||||| ||||||||||||||||||| |||||||||||||||||||||||||
Db   237 ATGGCTGCTCATTTTTTTCTGCTGTAGTGATTGGCTAGTGAAGGACCAGGATGAGAAGGTGG 178

Qy   816 ctggccgattatggtgacccgttaagttaggggaaggcttcaagtctcttagagccagggtg 875
         ||||||||||||||||||||| |||||||||||| |||||||||||||| ||||||||||
Db   177 CTGGCCAATTATGTGACCCGTAAGTTAGGGGAAGGCGTTCAAGTCTTTTAGAGCCAGGATG 118

Qy   876 gtactccgccatggccaagggccaagccatctacaattagtcaattgtctcatctcttaac 935
         ||||||||||||||| |||| |||||||||||||||||||| |||||||||||| |||||
Db   117 GTATTCTGCCATGGCCAAGGGCCAAGCCATTTCTACAATTAGTCGAGGGCCTATTCGTTAAC 58

Qy   936 aaaagaaccatatatattcctcaattcagcttaaggggcaacagcccccttacaaagtttct 992
         ||||||||||||||||||||||||||||||| |||||||||||||||||  ||||||||||
Db   57  AAAAGACCATATATATTCCTCAAATTCAGC-TTAAGGGCAACAGCCCCTTATAAGTTTCT 2

RESULT   6
AA435342
LOCUS       AA435342       426 bp    mRNA            EST        04-AUG-1997
DEFINITION  vd49g09.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
            IMAGE:803968 5' similar to WP:B0285.5 CE00644 ;, mRNA sequence.
ACCESSION   AA435342
VERSION     AA435342.1  GI:2140162
KEYWORDS    EST.
SOURCE      house mouse.
```

## Right column

```
ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1  (bases 1 to 426)
AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
           Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,B.,
           Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
           Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
           Waterston,R.
TITLE      The WashU-HHMI Mouse EST Project
JOURNAL    Unpublished (1996)
COMMENT    Contact: Marra M/Mouse EST Project
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: mouseest@watson.wustl.edu
           This clone is available royalty-free through LLNL ; contact the
           IMAGE Consortium (info@image.llnl.gov) for further information.
           MGI:480312
           Putative full length read
           Putative full length insert length is 463
           vector to vector length is 463
           Possible reversed clone: similarity on wrong strand
           Seq primer: -40m13 fwd.; ET from Amersham.

FEATURES   Location/Qualifiers
source     1..426
           /organism="Mus musculus"
           /strain="B6D2 F1/J"
           /db_xref="taxon:10090"
           /clone="IMAGE:803968"
           /clone_lib="Knowles Solter mouse 2 cell"
           /note="Organ: embryo; Vector: pBluescribe (modified);
           Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
           mRNA prepared from 13,500 2-cell stage embryos. Primer:
           SalI(dT): 5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTTT-3'. cDNAs
           were cloned into the MluI/SalI sites of a modified
           pBluescribe vector using commercial linkers (NEB).
           Average insert size:  1.2 kb."
           /dev_stage="2-cell"
           /lab_host="DH10B"
           /tissue_type="embryo"

BASE COUNT  126 a    97 c    107 g    96 t
ORIGIN

Query Match  91.1%;   Score 364.2;   DB 7;   Length 426;
Best Local Similarity  25.9%;   Pred No. 8.2e-97;
Matches  387;   Conservative  0;   Mismatches  38;   Indels  0;   Gaps  0;

Qy   589 tcatggagcaacagttacccgggacctggtcactgacctcaggaaagagagtgggtgttctttcc 648
         |||||||| ||||||||||| ||||||||||||||||||||||||| ||||||||||| || ||||
Db   2   TCATGGAGTACAGTTACCAGAGACCTGGTCACTGACCTCAGGAAAGGAGAGTGGGCCTTTCT 61

Qy   649 aacacaaaagctgtcaagccaacaacaagaataatgcccaagaagaagtggttggttgttgatggcg 708
         |||||||| |||||||| || |||||| || ||||||| |||| ||| |||| |||||||||||
Db   62  AACACAAAAAGCTGTCAAGCCAACCAACCAAATCATGCCCAAAAAGGTGGTTAGGTTGATTGCA 121

Qy   709 aaaggggaaggcttcctggacaacaattaccatctctaccacagcccaacatggctgctcc 768
         ||||| |||||| |||||||| || |||||||||||| |||||||||| |||||||| |||
Db   122 AAAGGGAAGGGAGGATTCCTGGACACACATTACCATCTCACCACAGCCCACATGGCTGCTTC 181

Qy   769 ttcgctgcccagtgactggctggtgaggaaccaggagatgagaaggcggctgcgattatg 828
         || ||||| |||||||||| ||||| |||| |||||||||||| || |||||| |||||||
Db   182 TTTGCTGCACAGTGACTGGCTGGTTGAGGAGGAACCAGGATGAGAAAGGTGGCTGCGCCAATTATG 241

Qy   829 gtgacccgtaagttaggggaaggcttcaagtctcttagagccaggggtggtactccgccatg 888
         ||||||| |||||||||| |||| ||| |||  ||||| ||||||| || |||||  ||||
Db   242 GTGACCCGGAAGTTAGGGAAGGGTTTAAATCTTTAGAACCAGGATGGTACTCGTCGCCATG 301

Qy   889 gcccaaggtcaagccatttctacacattagtcaggcctatctccttacaaaaaagaccatata 948
         ||||||| ||| |||||| || |||||||||| |||| |||| |||  || |||||||||||
Db   302 GCACAAGGGCAAGCCCATCCTCACCTTAGTCAYGCCCTATCTTCAGGGCGCCTAYGTA 361
```

RESULT 7

```
Qy    949 ttcctcaattcagtttaaggggcaacacgcccttacaagttctgtcaagaggcagcatgga 1008
          |||||||||||||||||||||||||||||||| ||||||||||||||||| ||||||||||
Db    362 TTCCTCAGTTCAGCTTTAAGGGCAACACGCCCATACAAGTTTCCGTCAGAGGCAGCATGGA 421

Qy   1009 gtcaa 1013
          || ||
Db    422 GTTAA 426
```

LOCUS           AA794925          455 bp    mRNA           EST      09-FEB-1998
DEFINITION      vr48g04.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone
                IMAGE:1123926 5', mRNA sequence.
ACCESSION       AA794925
VERSION         AA794925.1  GI:2857880
KEYWORDS        EST.
SOURCE          house mouse.
  ORGANISM      Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE       1  (bases 1 to 455)
  AUTHORS       Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
                Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
                Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
                Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R., and
                Waterston,R.
  TITLE         The WashU-HHMI Mouse EST Project
  JOURNAL       Unpublished (1996)
  COMMENT       Contact: Marra M/Mouse EST Project
                WashU-HHMI Mouse EST Project
                Washington University School of Medicine
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: mousesest@watson.wustl.edu
                This clone is available royalty-free through LLNL ; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                MGI:613262.
FEATURES             Location/Qualifiers
     source          1..455
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1123926"
                     /clone_lib="Knowles Solter mouse 2 cell"
                     /tissue_type="embryo"
                     /dev_stage="2-cell"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pBluescribe (modified);
                     Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
                     mRNA prepared from 13,500 2-cell stage embryos. Primer:
                     SalI(dT): 5'-CGGTCGACGTCGACCGTTTTTTTTTTTTTTTTT-3'; CDNAs
                     were cloned into the MluI/SalI sites of a modified
                     pBluescribe vector using commercial linkers (NEB)."
                     /note="Average insert size: 1.2 kb."
BASE COUNT      138 a     97 c     118 g     102 t
ORIGIN
```

Query Match     24.7%;  Score 346.6;  DB 12;  Length 455;
Best Local Similarity   88.0%;  Pred. No. 1.4e-91;
Matches   402;  Conservative   0;  Mismatches   49;  Indels   6;  Gaps   2;

```
Qy    594 gagcacagttacccgggaccctggtcactgacctcaggaagggaaggyagtgggtgcttctttccaaacac 653
          |||  ||||||||||||||||||||||||||||||||||||||||||  |||  |||| |||||||||||||
Db      5 GAGTACAGTTACCAGAGAGCTGGTCACTGACCTCAGGAAGGAGAGTGGGGCCTGTGTCTAACAC 64

Qy    654 aaaagctgtcaagccaacaacaagaataatgcccaagaaggtggtcaagatggtcttaggtgattgccgaaagg 713
          |||||||||  |||||||| |||||||  |||||||||||||||| ||||||||||||||||||||||| |||||
Db     65 AAAAGCTGTCAAGCCAACCAACAAAATCATGCCCAAAAAAGGTGGTTGGTTAGGTTGTCAATTGCAAAAGG 124
```

RESULT 8

```
Qy    714 gaagggcttccttgacaacattaccatctctacacacagcccacacatgggctgccttctttcgc 773
          ||||||| |||| ||||| ||||||| || ||| ||||||||||| ||||||||| |||||| |||
Db    125 GAAGGGATTCCTTGGACAACATTACCAATCTCAACCAACAGCACGCCCCACACATGGGCTGCCATTCTTTTGC 184

Qy    774 tgccagtgactggctggtgaggaaccaggagatgagaaagyggtgctgcgattatggtgac 833
          |||| ||||||||||||||||||||||||||||| |||||||| |||||||||||||||||
Db    185 TGCAAGTGACTGGCTGGTGAGGAACCAGGATGAGAAAGGTGGCGTGCTGGGCCAATTATGTGAC 244

Qy    834 ccgtaagttagggggaaggcttcaagtctttaagagccaagggtggtactccgccatggccca 893
          |||||||||||||||||||||||||||| ||| |||||||| ||||| ||||||||||||| ||
Db    245 CCGGAAGTTAGGGGGAAGGGGTTTAAATCTTTAGAACCAAGAGGTGGTACCTCTGCCATGGCCACA 304

Qy    894 agggcaagccattcctacacattagtcagggcctatctcttcttacaaaagaccatatattcct 953
          ||||||||||||||||||| |||||| ||||||| |||||||| |||||||||||||||| ||||
Db    305 AGGGCAAGCCATTCTCACCTTAGTCAGGGCCTATCTCTTCTTACAAAAGACCATATATTCCT 362

Qy    954 caattcagctttaagggcaacagcgcccctacaagttctgtcagaggcagcatgggagtcaa 1013
          ||||||||||||||||||||||| |||||||||||||||||||||| ||||||||| |||||
Db    363 TCAGTCAGCTTTAGGG----CACAGCCATACAAGTTTCCGTCAGAGACCAGCAGCATGGGAGTTAA 418

Qy   1014 ggctgtgttttatgaataaacactgactggtatgaagaa 1050
          ||| | |||| |||||||||||||| |||||||||||||
Db    419 AGCCGTGCTGTTCATGAATAAACATGACTGGTATGAGAGAA 455
```

LOCUS           CNS01WIY          866 bp    DNA            GSS      12-MAY-2000
DEFINITION      Tetraodon nigroviridis genome survey sequence T7 end of clone
                20UJ23 of library G from Tetraodon nigroviridis, genomic survey
                sequence.
ACCESSION       AL170467
VERSION         AL170467.1  GI:7808524
KEYWORDS        GSS; genome survey sequence.
SOURCE          Tetraodon nigroviridis
  ORGANISM      Tetraodon nigroviridis
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
                Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha;
                Holacanthopterygii; Acanthopterygii; Percomorpha;
                Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Tetraodon.
REFERENCE       1  (bases 1 to 866)
  AUTHORS       Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
                Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
                Weissenbach,J.
  TITLE         Charaterization and repeat analysis of the compact genome of the
                freshwater pufferfish Tetraodon nigroviridis
  JOURNAL       Unpublished
REFERENCE       2  (bases 1 to 866)
  AUTHORS       Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
                Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
                Saurin,W. and Weissenbach,J.
  TITLE         Human gene number estimate provided by genome wide analysis using
                Tetraodon nigroviridis DNA sequence
  JOURNAL       Unpublished
REFERENCE       3  (bases 1 to 866)
  AUTHORS       Genoscope.
  TITLE         Direct Submission
  JOURNAL       Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
                Genoscope.
  COMMENT       This sequence is a single read and was generated as part of a large
                scale clone-end sequencing project of the Tetraodon nigroviridis
                genome. For more information, please take a look at
                http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1..866
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="20UJ23"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG200CE12LP1-end : T7"
BASE COUNT      164 a    280 c    264 g    158 t
ORIGIN
```

```
RESULT   9

Query Match
Best Local Similarity   65.2%;     Score 331.6;  DB 230;  Length 866;
Matches  561;  Conservative  0;    Pred. No. 4.9e-87;
                                   Mismatches 224;  Indels  75;  Gaps   2;

Qy   466  aggtgtctgtggttctggaggacgacagaaaagaatcagcttcacttgtacacttagtc  525
          |||||  |  || |  |||| |||||||||| |||  ||||||  |  ||| |  |||| |  |||
Db     8  ACCGTGTCGGTCGTCCTCGGACAACACCGAGAAGGCCGCCGCCCTTGCTGATCCACTATGTG  67

Qy   526  tcaaataccagctaattgctcttaaagaaagagaagacaatatactatggcatcgggcccaga  585
          | |||  |  |||||  | |  |||||||  |||||||||||  |||  || ||  ||||| |||
Db    68  ACCAGCACGCAGCTGATAGCCTTCAAAGACCGGGAGGTCACTACCTACGGCATTGGTCCAAGG  127

Qy   586  acatcatggaagcacagttacccgggaccttggtcactgaacctcaggaaaggagtgggtctt  645
          |  ||  || |  ||||  |||  || |||| ||||||||  ||||||||||  | ||| | |||
Db   128  ACCGGCTGGAGCACCCTGACCAGGAGGACCTGCTCACTGACCTCAGGAAGGGCCGGTTGGCCTG  187

Qy   646  tccaacacaaaagctgtcaagccaacaaga-------------------  675
          ||||||||  |||  ||| ||||||||||||
Db   188  TCCAACACCAAGGCCGTGAAGGCCACGAAGTCAGTCGGCGTCCGGTTGGTTACGCAG  247

Qy   676  ------------------ataatgcccagaagg  691
                                |||||||||||| |||
Db   248  GTCGCGAGCGTCTGTGTAGAATCAGCGTTCCGATTGGTTTGTTCTGCAAGATCATGCCCAGCGCGGG  307

Qy   692  tggttaggttgattgcgaaaggaagggcttccttgacaacattaccatctctaccacag  751
          ||| | ||| |||| ||    | || ||  | || |||||| |||||||| |||||||||
Db   308  TGGTCGCGGCGTGGTGCTACACGGGCGGGGGGGGCCGTTCGTTGACAACGTCACCATCTCCACCACTG  367

Qy   752  cccacatggctgccttcttccgctgcgtctggtcgggtgaggaaccaaggatgagaaag  811
          |||||||||||||||||||||| ||||||||  ||  |||  ||||| ||||||||||||
Db   368  CCCACATGGCTGCCTTCTTCCGCCTGCGTCTGCGCCCCAATCAGGACGAGGACGGGTG  427

Qy   812  gcgggctggccgattatggtggacccgtaagttaggggaggcttcaagtcttagagccag  871
          ||||| |||||| | ||||| |  ||||  | |||||  ||   ||  ||  |||||||||
Db   428  GCGGGTTGGCCCCATCATGGTCACCCGCCCAAACTAGGCGTTCCGGGCCCTGGAGCCCG  487

Qy   872  ggtgtgttactccgccgccatggggccaaagggccattctgaacattaagtcaggggcctatctct  931
          |||| | ||||||||||||||| |||   |||||||| |||||| |||||  |||||||||||||||
Db   488  GCTGGTACTCGGCCATGGCCTCAGGGTGTCAGGCCATGTCACCCTGGTGAGAGCCTACCTCT  547

Qy   932  taacaaaagaccatatattcctcaattcagcttaaaggcaacagcccccttacaagttc  991
          | ||||||||||  | |||  ||||||| ||  ||||| ||||||||||  ||||| |||
Db   548  TCACCAAAAGACCAGGCTTTATCTCAACGCCCCGGTCGGACCCCTACAAGGTGC  607

Qy   992  tgtcaagcagcagatggagtcaaggctgtgttgttttatgaataaacatgactggttatgagaagaat  1051
          | ||||||||| | | |||||||  ||| |||  | |||  ||||| || ||||||| ||||| |||
Db   608  CTTCAGCAGCACAGCAGCGGGGTCAAAGCCCGTGTTCATGAACAAGTACTACGACTGGTTACGAGGAGT  667

Qy  1052  atccaactacaacctagctccttttgttttaaatggcttatgtatctttaattgggctgt  1111
          || |||| |||| ||| ||||||| ||| | ||||| || || |||| || ||||||| |
Db   668  ACCCCAACCACACCAGCTCCTTCGTGTCTACTCCCTCCTGCTGGGGCTCT  727

Qy  1112  atggctaaaagaaactcgagggaaaagaagcgcagggtccttgtatgagc  1171
          | |||  ||||||||| |||| ||||||| | |||||  || || ||| |
Db   728  TTGACCTGACCGAGACGGGAGCGGGGAGGGCCGCGAGCGTTCTTCAG-CC  786

Qy  1172  gtggcatggaatccctaaagccatgcatgctcccttgtacgacactggctcaaggaaccatct  1231
          | ||||||||| ||||| ||||||||| || | ||| ||||| || |||||  | ||| ||||
Db   787  ACGGCATGGAGTCCCTGAAAGGCCATGCTGCCGTCTTCGACACGGGCTCCGGGGCCGCATCT  846

Qy  1232  atgacctccggcacttcatg  1251
          |||||||  ||||||||||||
Db   847  ACGACCTGCGCCACTTCATG  866
```

```
Query Match
Best Local Similarity   90.8%;     Score 304.2;  DB 20;  Length 358;
Matches  324;  Conservative  0;    Pred. No. 4.8e-79;
                                   Mismatches  33;  Indels   0;  Gaps   0;

Qy   582  cagaacatcatggagcacagttacccgggacctggtcactgacctcaggaaggaggtggg  641
          ||||||  ||||||||  ||||||||| | |||||||||||||||||||||||||   |||
Db     1  CAGAACTTCATGGAGTACAGTTACCACAGAGACCTGGTCACTGACCTCAGGAAGGGTTAGGT  60

Qy   642  tctttccaacacacaaagctgtcaagccaacaagaataatgcccaagaaggtggttaggtt  701
          ||||||| ||||| ||||||||||||||||||||| |||| |||||| ||||||||||||||
Db    61  CCTTTCTAACACAAAAAGCTGTCAAGCCAACAAAATCATTGCCCAAAAGGTTGGTTAGGTT  120

Qy   702  gattgcgaaaggagaaggcttccttgacaacattaccatctctaccacagcccacatggcc  761
          |||||| |||||| || ||  ||||||| |||| |||||||| |||| |||||||||||| |
Db   121  GATTGCAAAAGGGAGAGGATTCCTTGGCAACAATTACCATTACCACCACAGCCCACATGGC  180

Qy   762  tgcttcttcgctcgacctggctggctggtgaggaaccaggagagaaaggcgggtggttgactc  821
          ||| | ||| |||  ||||||| | | || |||||||| |||| ||||| | ||   || |||
Db   181  TGCATTCTTCTTCGGCTCGGACCTGGCTAGTGAGGAACCCAGGATGAGAAAAGGTTGGGCTCGGCC  240

Qy   822  gattatggtgacccgtaagttaggggaggcttcaagtcttagagccaggggtggttactc  881
          |||||||||||||||  |||||||||||| |  |||  |||||| |||||| |||| |||
Db   241  AATTATGGTGACCCGGAAGTTAGGGGAGGGCGGTTTAAATCTTTAGAACCAGGATGGTACTC  300
```

RESULT 10

```
Qy    882 cgccatggtggcccaagggcaagccaagttctacattagtcaggggccctatctcttaacaa 938
              ||||||| |||| |||||||||||||||||||||||||||||||||||||| ||||| ||
Db    301 TGCCATGGCACAAGGGCAAGCCAGCCATCTACCTTAGTTCAGGGCCTATCTTCTAACGAA 357
```

```
LOCUS       BE170776                  434 bp    mRNA              EST     21-JUN-2000
BE170776/c
DEFINITION  QV4-HT0539-160300-137-e07 HT0539 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE170776
VERSION     BE170776.1  GI:8633497
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=t2=QV4-HT0539-160
            300-137-e07&t3=2000-03-16&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 434.
FEATURES             Location/Qualifiers
     source          1..434
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0539"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      122 a      87 c      87 g     138 t
ORIGIN
```

Query Match    21.6%;   Score 303;   DB 163;   Length 434;
Best Local Similarity   91.5%;   Pred. No. 1.2e-78;
Matches   321;   Conservative   0;   Mismatches   30;   Indels   0;   Gaps   0;

```
Qy    360 tgttaaacagttcattgctccagaaaccagtgaaggtgtatccttgcaactggggaacac 419
              ||||| ||||||||||||||||||||| || |||||| ||||||| ||||||||||||
Db    363 TGTTGGTATAATTCTCCCCTACAGGAACCACTGAAGGTGTATCCTTGCAACTGGGAAACAC 304
```

```
Qy    420 aaaagatttattattattcattgacctcaagtcttaacacaaatggaagcgtgtctgtggt 479
              ||||||||||||||||||||||| ||||||| ||  |||||| |||||| | |||||
Db    303 AAAAGATTTATTTATTCATTGACCTCAAGTTCTTGACAAAATGGAAGTGTGTCCGTGGT 244
```

```
Qy    480 tctggagacgacgaaaagaatcagctcagtctatcctacattatgtctcaaaataccagct 539
              |||||| |||||||||| |||| ||||||||||| |||||||| |||||||  ||| ||
Db    243 TCTAGAGACCACAGAAAAGAATCAGCTCTTCACTATACATTATATGTCTCAAATGCTCAGCT 184
```

---

RESULT 11

```
Qy    540 aattgctttaaagaaagaagagacaatatcatgtggtcatcgggccccaagaacatcatggaagcac 599
              ||||||| || |||||||||||||||||||||||| ||||||||| || |||| ||||||||||||
Db    183 AATTGCTTTTAAAGAAAGAAGAGATATATTAACTATATGGCATTGGGCCCAGGAAAACTTCATGGAGCAC 124
```

```
LOCUS       BF080558                  566 bp    mRNA              EST     18-OCT-2000
BF080558
DEFINITION  231598 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF080558
VERSION     BF080558.1  GI:10874460
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 566)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 40  row: B  column: 6
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..566
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      182 a     107 c     135 g     142 t
ORIGIN
```

Query Match    20.3%;   Score 285.4;   DB 143;   Length 566;
Best Local Similarity   94.9%;   Pred. No. 2.1e-73;
Matches   295;   Conservative   0;   Mismatches   16;   Indels   0;   Gaps   0;

```
Qy     13 tctcatagctattccaaagtctattgcacagagaagccccttatcaccctgatggtgtgtt 72
              || |||||||||||||||||||||  |||||||| ||||||||||||| ||||||||||||
Db    256 TCTCATAGCTATTCCAAAGTCTATACACAGAAGACGCCCTTATCACCCTGATGGTGTGTTT 315
```

```
Qy     73 atgtcccttgaaggctacaaatgtgaaagtccgagacagagtcaagtgcataagtggggtt 132
              ||||||||||||||||||||||||||||||||||||| |||||||||||||| |||||||||
Db    316 ATGTCCCTTGAAGGCTACAAATGTGAAGTCCGAGACAGAGTCAAATGCATAAGTGGGGTT 375
```

```
Qy    133 gaaggtgtaccttatctacacagtggggacctcaaggctattctaccaaatccagatt 192
              ||| ||||||||||  |||||||||||||||||||||||| ||||||||||||||||||
Db    133 ...
```

```
Db   376 GAAGGTGTAACCTTTGTCTACACAATGGGACCTCAAGGCTATTTCTACCAATCCAGATT
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   313 aagggctgctt 323
         |||||||||||
Db   556 AAGGGCTGCTT 566
```

RESULT 12
AI981391
LOCUS      AI981391       590 bp    mRNA                    EST      10-JUL-2000
DEFINITION pat.pk0052.c6.f chicken activated T cell cDNA Gallus gallus cDNA
           clone pat.pk0052.c6.f 5' similar to KIAA0836 protein, mRNA
           sequence.
ACCESSION  AI981391
VERSION    AI981391.1   GI:5884419
KEYWORDS   EST.
SOURCE     chicken.
ORGANISM   Gallus gallus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
           Phasianinae; Gallus.
REFERENCE  1  (bases 1 to 590)
AUTHORS    Tirunagaru,V.G., Sofer,L., Cui,J., and Burnside,J.
TITLE      An expressed sequence tag database of T-cell-enriched activated
           chicken splenocytes: sequence analysis of 5251 clones
JOURNAL    Genomics 66 (2), 144-151 (2000)
MEDLINE    20318616
COMMENT    Contact: Joan Burnside
           Molecular Endocrinology
           University of Delaware
           40 Townsend Hall, Newark, DE 19717, USA
           Tel: 302 831-1345
           Fax: 302 831-3651
           Email: Joan@Udel.Edu
           Seq primer: T7
FEATURES            Location/Qualifiers
     source         1..590
                    /organism="Gallus gallus"
                    /db_xref="taxon:9031"
                    /clone="pat.pk0052.c6.f"
                    /clone_lib="chicken activated T cell cDNA"
                    /sex="male"
                    /cell_type="Con A-activated splenic T cell"
                    /lab_host="E.coli TOP10 F'"
                    /note="Vector: pcDNA3"
BASE COUNT    146 a    165 c    160 g    113 t    6 others
ORIGIN

Query Match    20.3%;    Score 285.4;    DB 104;    Length 590;
Best Local Similarity    76.0%;    Pred. No. 2.2e-73;
Matches    352;    Conservative    0;    Mismatches 111;    Indels    0;    Gaps    0;

```
Qy   119 gcataagtgggttgaagtgtacctttatctacacagtgggacctcaaggctattct 178
         |||||||| ||||||||||| |||||||||||||||||||||||||||||||||||
Db     1 GCATAAGTGGTGTTGAAGGTGTGCCCTTATCCACCACAGTGGGGACCTCAGGGCTACTTCT 60
Qy   179 acccaatccagattcgacagtatgggttaagtcactacacagcagaatcaactgaaaac 238
         ||||||| ||||||| ||||||||||||||| |||||||||||||||||||||||| ||
Db    61 ATCCCAATCCAGATCCGCACAGTACGGGTGAGTCACTACACAGCAAAAAACCTTGACCGAGAAGC 120
Qy   239 ccccttccatatagagtgtatgaaacgacagggacacaaaacagcgcaagcccaatg 298
         ||||| ||||||| |||||| ||||| |||||| |||| | |||||||||| |||||
Db   121 CCCCGCACATCGAGTGTGTACGAAAACGGCGCCGAGAGAGGCACAGGGGCACAGGGCCTGCTG 180
```

RESULT 13
BE861166
LOCUS      BE861166       381 bp    mRNA                    EST      29-SEP-2000
DEFINITION UI-M-AM0-adu-h-07-0-UI.r1 NIH_BMAP_MAM Mus musculus cDNA clone
           UI-M-AM0-adu-h-07-0-UI 5', mRNA sequence.
ACCESSION  BE861166
VERSION    BE861166.1   GI:10378842
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1  (bases 1 to 381)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
           discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
COMMENT    Contact: Chin, H
           National Institute of Mental Health
           6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
           20892-9643, USA
           Tel: 301 443 1706
           Fax: 301 443 9890
           Email: mEST@mail.nih.gov
           cDNA Library Preparation: M.B. Soares Lab Clone distribution:
           Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
           should be noted that Bento Soares is generating a small number of
           additional specialized non-redundant arrays of BMAP cDNAs whose
           availability will be considered under appropriate and limited
           collaborative arrangements
           Seq primer: M13 Reverse.
FEATURES            Location/Qualifiers
     source         1..381
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="taxon:10090"
                    /clone="UI-M-AM0-adu-h-07-0-UI"
                    /clone_lib="NIH_BMAP_MAM"
                    /dev_stage="27-32 days"
                    /lab_host="DH10B (life Technologies)"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; The
                    NIH_BMAP_MAM library is a non-normalized library
                    constructed from mouse amygdala. The tag is a string of 5
                    nucleotides present between the Not I site and the
                    oligo-dT track. The library was constructed as described
                    by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
                    , 1996. Tissue provided by Ms. Annie Novakovich,
                    Zivic-Miller Laboratories."
BASE COUNT    100 a    110 c     82 g     89 t

```
Qy   193 gccacagtgggttaagtcactcacacagccaagaatcaactgaaaaccccctcatatagag 252
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   436 GCACAGTATGGATTAAGTCATTCATTACACTAAGAATCTAACTGAGAAAACCCCTCACAATAGAA 495
Qy   253 gtatatgaaacacagcagaagacaggaacacaacagcccaatgactggactgtgtgcc 312
         |||||||| ||||| |||||| |||| ||||| ||||||||||||||||||||||
Db   496 GTATATGAAACACACAGCAGAAGACAGGGACAAAAAAGTAACAAGCTCAGTGACTGGACTGTGCCA 555

Qy   299 actgggactgtgtgcccaaagggctgctttatggtagtgtggggctgataagtcaagattcaacca 358
         |||||| |||||||||||||| |||| ||||| |||||||| | |||||| |||||||||||
Db   181 AGTGGCACGGTGTCGCCCAAAGGCTGCTCCCTTTCCACGGTGCCGGATAAAGCCAAGTTCACCA 240
Qy   359 atgttaaacagttcatcgtccagagaaaccagtgaaggtgtatcctgcaactgggaaca 418
         |||||||| ||||||||||||||| |||||||||||||||||||| |||||||||||
Db   241 GTGTTAAACACTTTGTTGGTCCAGAGAACACCGAGGGGTTTCCCTGACCAGCTTGGCAACG 300
Qy   419 caaaagattttattattcattgacctcaagtcttcataaaatggaaggtgtctgtgg 478
         ||||||||||||||||| || ||||||||| |||||| |||||||||||| |||||
Db   301 CCACAGATTTCATTATTTTCTTTCGATCTCAAGCTCGTAACCAACGGACCAGTTCTGTGTGG 360
Qy   479 ttctggagacgaagaaagagacaatactacatatggcatcgggcc 581
         |||||||||||||| | |||| |||| ||||||||||||| |||
Db   361 TGCTTGGAGACCACCGGAGAAGAAACCACTGCTTCACCGTGCACTACGGATCGGGGGCC 420
Qy   539 taattgctctttaaagaagaagacaatatactatggcatcgggcc 581
         || ||| ||| || ||||| ||||| |||||||||||||| |||
Db   421 TCATCGCCGCTTCAGGGAGCGGCGACATCTPACTACGGATCGGGGGGCC 463
```

ORIGIN

RESULT 14

Query Match
Best Local Similarity   93.2%;
Matches 204;   Conservative   0;   Pred. No. 1.2e-46;   Length 381;   Indels 0;   Gaps 0;
Mismatches 15;

```
Qy  1186 cttaaagccatgctccccttgtacgacactggctcaggagaccactctatgacctccggac 1245
         ||||||||||||||||||||||||||||||||| | |||||||||||||||||||| |||||
Db     1 CTTAAAGCCATGCTGCCCTTGTATGATACTGGCTCCGGGGACCATCTATGACCTCCGGCAC 60

Qy  1246 ttcatgcttggcattgccccccaacctggtcccgcgctgggggactatacacacccacc 1305
         ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    61 TTCATGCTTGGCATTGCTCCCCAACCTGGGCCGCGCTGGGGACTATCACACACCCACC 120

Qy  1306 caactgcagctgcttagcaccattgatgagtccccaatcttcaaagaatttgtcaaagg 1365
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CAGCTCCAGCTGCTCAGCACCATCGATGAGTCCCCAAATCTTCAAAGAAATTTGTCAAAGAGG 180

Qy  1366 tggaagagctaacttaaaggcagccggggcaaagcacaac 1404
         ||||||||||||||||||||||||||||||||||||||||
Db   181 TGGAAAAGCTACCTTAAAGGCAGTAGGGCAAAAGCACAC 219
```

LOCUS       BB579985          251 bp    mRNA           EST       30-NOV-2000
DEFINITION  BB579985 RIKEN full-length enriched, in vitro fertilized eggs Mus
            musculus cDNA clone 7420402F02 5', mRNA sequence.
ACCESSION   BB579985
VERSION     BB579985.1  GI:11475932
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 251)
  AUTHORS   Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P.
            , Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodoyama,Y.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno
            ,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakanura,M., Nishi,K.,
            Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,
            Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A.,
            Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
            ,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
  JOURNAL   Unpublished (2000)
  COMMENT   Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            Carninci,P.: Nishiyama,Y.: Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            URL:http://genome.rtc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.

FEATURES             Location/Qualifiers
     source          1..251
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="7420402F02"
                     /clone_lib="RIKEN full-length enriched, in vitro
                     fertilized eggs"
                     /sex="female"
                     /tissue_type="in vitro fertilized eggs"
                     /dev_stage="egg"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5',
                     GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTVN 3'], CDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5',
                     GAGAGAGAGAGTTCTCGAGTTAATTAAATTAAATTAAATCCCCCCCCCCCCCC 3']. cDNA
                     was cloned into the XhoI and BamHI sites. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT     71 a     36 c     67 g     77 t
ORIGIN

RESULT 15

LOCUS       AV672171          684 bp    mRNA           EST       05-OCT-2000
DEFINITION  AV672171 Nori Satoh unpublished cDNA library Ciona intestinalis
            cDNA clone cit2b15 5', mRNA sequence.
ACCESSION   AV672171
VERSION     AV672171.1  GI:10110170
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 684)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-1,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)

Query Match
Best Local Similarity   78.5%;
Matches 197;   Conservative   0;   Pred. No. 1.1e-37;   Length 251;   Indels 0;   Gaps 0;
Mismatches 54;

```
Qy   34 tatgcacagagagcccctatcacccttgaatggtgtgttatgtcctttgaaggctacaat 93
        ||||||||||| ||| | |||||||||| || |||||| || ||||| | || |||
Db    1 TATGCACAGAGAAATTACCTTATCACCCTGATCGGTGTGTTTATGTCTTGTTGTTTGAACGGTACAAT 60

Qy   94 gtgggaagtccgagacagaggtcaagtgcataagtggggttgaaggtgtaggtgtacctttatctaca 153
        |||||||| ||| ||| |||||| |||| | | |||||||| ||| ||||| | | ||| ||
Db   61 GTGGGAAATCCGATAAAGAGCTCGAATGTATAAGTGGAATTGAAGGTGTGCCATTATCTTCC 120

Qy  154 cagtggggacctcaaggctatttctacccaatccagattgcacagtatgggttaagtcac 213
        |||| || |||| ||||| || ||| ||||| ||| ||| || ||||| ||||| | ||
Db  121 CAGTGGGGGCCTCTAGGGTATTTCTACCGAAATCCGTATTGCACAGTATGGGGTAAGTCAT 180

Qy  214 tacagcaggaatctaactgaaaaacccctcatatagaggtatatacaaacagcagaaagac 273
        |||| | |||| | ||| |||||||| || | ||||||||||| ||| |||| ||||||
Db  181 TACAACTAGTAGAATTTTTCCGAGAAACCGCCTAGGAGGTTGAGTTGATGATATAGGGAATTAG 240

Qy  274 aggggacaaaaa 284
        ||| | || |
Db  241 TGGGGTATATA 251
```

```
COMMENT      Contact: Nori Satoh
             Department of Zoology
             Kyoto University
             Sakyo-ku, Kyoto 606-8502, Japan
             Tel: 81-75-753-4081
             Fax: 81-75-705-1113
             Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source  1. .684
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="citb2b15"
                     /clone_lib="Nori Satoh unpublished cDNA library"
                     /tissue_type="whole animal"
                     /dev_stage="tailbud"
BASE COUNT      230 a      113 c      145 g      196 t
ORIGIN

Query Match          10.2%;   Score 143.6;   DB 31;   Length 684;
Best Local Similarity   52.1%;   Pred. No. 2.6e-31;
Matches  345;   Conservative  0;   Mismatches  314;   Indels   3;   Gaps   1;
```

```
Qy    165 tcaaggctatttctaccaatccagattgcacagtatgggttaagtcactacagcaagaa 224
              |||||| ||  ||    ||||||||||||||||| || ||||||| |||||||||||||
Db      4 TGAAGGTTACTATTATGGAATCCAGANTGCTCAGTTTGGCATGAGTTGCTACTACACCAAACA 63

Qy    225 tctaactgaaaaaccccctcatatagaggtatatgaaacagcagaagacagggacaaaa 284
              || | |||| | |||| ||  | ||| |||| | ||  || ||||| | |   ||  +
Db     64 CATGACACAAAAAGCCACCAACCAAACAAGTTTATGAAAATGCAGAGGGTTCTGTCCAGTC 123

Qy    285 cagcaagcccaatgactggactgtgccaagggcgtcttatgggtcagtgtgggctgataa 344
              |||| ||                            || |||||||| ||
Db    124 TCGGTGGTCCTCCAGTAATAAATACTGTTTGGTTGAAAACGAAGAGGATAAAGAAAAGGAA 183

Qy    345 gtcaagattcaccaatgttaaacagttcattgctccagaaaccagtgaaggtgtatcctt 404
              || |   |||||||||  ||||  ||   ||||| |||  ||| || ||| |||   ||| |
Db    184 TTCTCGAGTTATTAAATTCAGACTCCATTGCTATTCCATTGGGAAGTGGGCGCCTCCCT 243

Qy    405 gcaactgggaacacaaagatttattattttcattttgacctcaagttcttaacaaatgg 464
              ||  |                |||||||||| |   || |||||||||||||||||| ||
Db    244 TGTTCTTGGCACACCANTGGAGTTTATCTTGTCGTTTGATTTAAAATGGTTGCAAATGG 303

Qy    465 aagcgttgtctgtggttctggagacgacaagaaaaagaaatcagctcctcactgtacattatgt 524
              ||    ||  ||||||| |      || |    ||| | | ||| ||||| ||| |  || | |
Db    304 TTCCTTATCTGTTGTTCTTGGAAACCAATGACAAAAATGACAAAATAATACAAATTCATTTACAT 363

Qy    525 ctcaaataccacagctaattgtctttaa---agaaagagacatatactatatggcatcgggcc 581
              |||||| |  || ||| | || |||        | |||| || | || ||||| |||||||||||
Db    364 CACAAAACAGTTTGGTCATTGATTTATGATTTGAAATCAAATATATATTATTACGGAATCGGTCC 423

Qy    582 cagaacatcaatggagcacaagttaccccgggacctggtcactgacctcaggaaaggagtgggg 641
              ||| |                       ||| ||  |||||  ||| ||||| |||||| ||||
Db    424 AAGTCGTACTTGGAGACATATGGCAAGGGCGATTTTGTTGCAGATTTGAAAAAAGGTATTGG 483

Qy    642 tctttcccaacaacaaaagctgtcaagccaacaaggaataatgcccaagaagtggttaggtt 701
              ||||| ||  |  ||      || || | |  |||  |    ||  |||| ||||| ||||| |
Db    484 TCTTACAACTAAAAACAAGCAANAGAAAGTTAGAAATTGCCCATTCAAAAGGTTTCAAGGCT 543

Qy    702 gattgcgaaaagggaagggctttccttgacaacatctacccacagcccacatggc 761
              | ||||  |   | |||  |  |   || || || || || ||| ||||| | |  |
Db    544 TATTGTNAAGGGTCCTGGTTATATTGATAATGTGACACTCTCCACGTTCTGAACACTTGAC 603

Qy    762 tgccttcttcgctgccagtgactggctggtgaggaacacaggagcgggcggccgcc 821
              ||||| |  ||||||| ||||||| |||| ||  |||| |||||||  |  ||   ||
Db    604 CCAATTTTTTTGATGCAAGCCAATGCATGTATTTAAACCAAGACAAGAGACTGGAGGGTG 663

Qy    822 ga 823
              ||
Db    664 GA 665
```
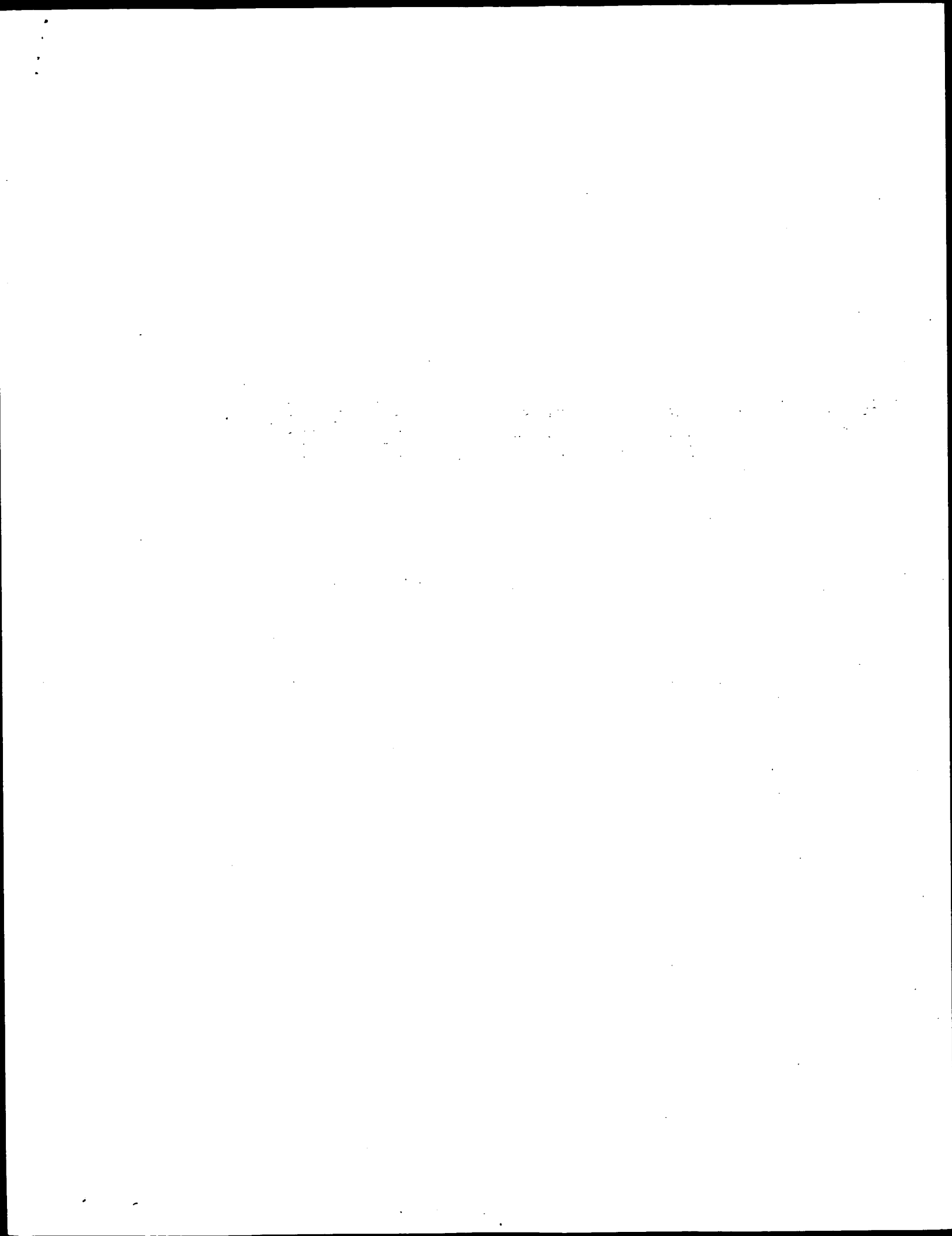
OM nucleic - nucleic search, using sw model

Run on:         May 25, 2001, 02:20:59 ; Search sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Title:          US-09-403-269-12_COPY_1_1404
Perfect score:  1404
Sequence:       1 tccaagctgaattctcaatag.........gcagccgggcaaagcacaac 1404
                                             (without alignments)
                                             Search time 1798.43 Seconds
                                             11513.372 Million cell updates/sec
Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       1283235 seqs, 737929652 residues

Total number of hits satisfying chosen parameters:     2566470

Post-processing: Minimum Match 0%
                 Maximum Match 100%
Minimum DB seq length: 0
Maximum DB seq length: 200000000
                 Listing first 45 summaries

Database :      GenEmbl:*
          1:    gb_ba1:*
          2:    gb_ba2:*
          3:    gb_ba3:*
          4:    gb_in1:*
          5:    gb_in2:*
          6:    gb_in3:*
          7:    gb_om:*
          8:    gb_ov:*
          9:    gb_pat1:*
         10:    gb_pat2:*
         11:    gb_ph:*
         12:    gb_pl1:*
         13:    gb_pl2:*
         14:    gb_pl13:*
         15:    gb_pl4:*
         16:    gb_ba1:*
         17:    gb_ba2:*
         18:    em_fun:*
         19:    em_htgo_hum:*
         20:    em_htgo_inv:*
         21:    em_htgo_rod:*
         22:    em_htg_hum1:*
         23:    em_htg_hum2:*
         24:    em_htg_hum3:*
         25:    em_htg_hum4:*
         26:    em_htg_hum5:*
         27:    em_htg_hum6:*
         28:    em_htg_hum7:*
         29:    em_htg_hum8:*
         30:    em_htg_inv1:*
         31:    em_htg_inv2:*
         32:    em_htg_other:*
         33:    em_htg_rod:*
         34:    em_hum1:*
         35:    em_hum2:*
         36:    em_hum3:*
         37:    em_hum4:*
         38:    em_hum5:*
         39:    em_hum6:*
         40:    em_hum7:*
         41:    em_in:*
         42:    em_om:*
         43:    em_or:*
         44:    em_ov:*
         45:    em_pat:*
         46:    em_ph:*
         47:    em_pl:*
         48:    em_ro:*
         49:    em_sts:*
         50:    em_sy:*
         51:    em_un:*
         52:    em_vi:*
         53:    gb_sts1:*
         54:    gb_sts2:*
         55:    gb_sts3:*
         56:    gb_sy:*
         57:    gb_un:*
         58:    gb_vi1:*
         59:    gb_vi2:*
         60:    gb_htg1:*
         61:    gb_htg2:*
         62:    gb_htg3:*
         63:    gb_htg4:*
         64:    gb_htg5:*
         65:    gb_htg6:*
         66:    gb_htg7:*
         67:    gb_htg8:*
         68:    gb_htg9:*
         69:    gb_htg10:*
         70:    gb_htg11:*
         71:    gb_htg12:*
         72:    gb_htg13:*
         73:    gb_htg14:*
         74:    gb_htg15:*
         75:    gb_htg16:*
         76:    gb_htg17:*
         77:    gb_htg18:*
         78:    gb_htg19:*
         79:    gb_htg20:*
         80:    gb_htg21:*
         81:    gb_htg22:*
         82:    gb_htg23:*
         83:    gb_htg24:*
         84:    gb_htg25:*
         85:    gb_pr1:*
         86:    gb_pr2:*
         87:    gb_pr3:*
         88:    gb_pr4:*
         89:    gb_pr5:*
         90:    gb_pr6:*
         91:    gb_pr7:*
         92:    gb_pr8:*
         93:    gb_pr9:*
         94:    gb_ro1:*
         95:    gb_ro2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 1404 | 100.0 | 3085 | 7 | AF003927 | AF003927 Bos tauru |
| 2 | 1249.6 | 89.0 | 2121 | 89 | AK002188 | AK002188 Homo sapi |
| 3 | 1249.6 | 89.0 | 4791 | 85 | AB020643 | AB020643 Homo sapi |
| 4 | 1203.2 | 85.7 | 2088 | 94 | AF325532 | AF325532 Mus muscu |
| 5 | 933 | 66.5 | 167187 | 71 | AC026992 | AC026992 Homo sapi |
| 6 | 490.6 | 34.9 | 171073 | 75 | AC073380 | AC073380 Homo sapi |
| 7 | 204 | 14.5 | 167187 | 71 | AC026992 | AC026992 Homo sapi |
| c 8 | 150.8 | 10.7 | 29372 | 65 | AC018117 | AC018117 Drosophil |
| c 9 | 150.8 | 10.7 | 119557 | 60 | AC007593 | AC007593 Drosophil |
| c 10 | 150.8 | 10.7 | 167062 | 60 | AC007624 | AC007624 Drosophil |
| c 11 | 150.8 | 10.7 | 299474 | 4 | AE003789 | AE003789 Drosophil |

RESULT 1
AF003927

| # | Score | | | Accession | Description |
|---|------|-----|-----|-----------|-------------|
| 12 | 75.8 | 5.4 | 41397 5 | CEB0285 | 234533 Caenorhabdi |
| 13 | 42.2 | 3.0 | 156257 79 | AL160170 | Homo sapi |
| 14 | 42.2 | 3.0 | 16721 89 | AL158841 | Human DNA |
| 15 | 41.6 | 3.0 | 174788 83 | CNS01DVU | Homo sapi |
| 16 | 41.4 | 3.0 | 18758 91 | HS45QC20 | Human DNA s |
| 17 | 41.4 | 2.9 | 16971 79 | Z84720 | Human DNA s |
| 18 | 40.6 | 2.9 | 159705 66 | AL157778 | Homo sapi |
| 19 | 40.6 | 2.9 | 159705 74 | AC021007 | Homo sapi |
| 20 | 40.6 | 2.9 | 174499 74 | AC068694 | Homo sapi |
| 21 | 40.6 | 2.9 | 175179 74 | AC072025 | Homo sapi |
| 22 | 40 | 2.9 | 217257 67 | AC022681 | Homo sapi |
| 23 | 39.6 | 2.8 | 213266 74 | AC069063 | Homo sapi |
| 24 | 38.6 | 2.8 | 15456 10 | AX073600 | Sequence |
| 25 | 38.6 | 2.8 | 15456 58 | AF178654 | Bovine pa |
| 26 | 38.6 | 2.7 | 42461 77 | AC083927 | Homo sapi |
| 27 | 38.6 | 2.7 | 311757 67 | AC022222 | Homo sapi |
| 28 | 38.4 | 2.7 | 2738 14 | Z73195 | S.cerevisia |
| 29 | 38.4 | 2.7 | 40383 15 | SCYLR023C | SCYLR023C |
| 30 | 38.4 | 2.7 | 160466 76 | SPCC417 | S.pombe c |
| 31 | 38.4 | 2.7 | 214780 74 | AC079386 | Homo sapi |
| 32 | 38.4 | 2.7 | 242662 77 | AC069075 | Homo sapi |
| 33 | 38.2 | 2.7 | 85329 90 | AC080020 | Mus muscu |
| 34 | 38.2 | 2.7 | 150351 94 | AL445223 | Human DNA |
| 35 | 38.2 | 2.7 | 156823 67 | AC074047 | Mus muscu |
| 36 | 38.2 | 2.7 | 159000 65 | AC022051 | Homo sapi |
| 37 | 38 | 2.7 | 210974 75 | AC018427 | Mus muscu |
| 38 | 37.6 | 2.7 | 42412 12 | AC074048 | Neurospor |
| 39 | 37.6 | 2.7 | 43556 12 | AC067918 | Neurospor |
| 40 | 37.6 | 2.7 | 102990 82 | AC067937 | Neurospor |
| 41 | 37.6 | 2.7 | 109568 82 | AP000801 | Homo sapi |
| 42 | 37.6 | 2.7 | 172024 76 | AP000840 | Homo sapi |
| 43 | 37.6 | 2.7 | 175885 82 | AC078971 | Homo sapi |
| 44 | 37.6 | 2.7 | 183615 82 | AP001132 | Homo sapi |
| 45 | 37.6 | 2.7 | 187987 61 | AP001494 | Homo sapi |
| | | | | AC009639 | Homo sapi |

ALIGNMENTS

LOCUS       AF003927      3085 bp    mRNA    MAM    04-OCT-1997
DEFINITION  Bos taurus C5-glucuronyl epimerase mRNA, partial cds.
ACCESSION   AF003927
VERSION     AF003927.1  GI:2465198
KEYWORDS    .
SOURCE      cow.
ORGANISM    Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 3085)
  AUTHORS   Li,J.-P., Hagner-McWhirter,E., Kjellen,L., Palgi,J., Jalkanen,M.
            and Lindahl,U.
  TITLE     cDNA cloning and expression of D-glucuronyl C5-epimerase from
            bovine lung
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3085)
  AUTHORS   Li,J.-P. and Lindahl,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1997) Medical and Physiological Chemistry,
            Biomedical Center, Husargatan 3, Uppsala 751 23, Sweden
FEATURES             Location/Qualifiers
     source          1..3085
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /tissue_type="lung"
     CDS             <73..147
                     /codon_start=1
                     /function="heparin/heparan sulfate biosynthesis"
                     /product="C5-glucuronyl epimerase"
                     /protein_id="AAB72083.1"
                     /db_xref="GI:2465199"
                     /translation="MSFGYNVEVRDRVKCISGVEGVYPLSTOWGPQGYFYPIQAOYG
                     LSHYSKNLIEKPPHIEVVETAEDRDKNSKPNDMTVPKGCFMASVADKSRFNVKQFIA
                     PETSEGVSLOGNTKDFIISFDLKFLTNGSVSVVLETTEKNQLFTVHYVSNTOLIAFK
                     ERDIYYGIGPRTSMSTVRDLVTDLRKGVGLSNTKAVKPTRIMPKKVRLIAKGKGFL
                     DNIITSTTAHMAAFFAASDWLVRNQDEKGGWPIMVTRKLGEGFKSLEPGWYSAMAQGO
                     AISTLVRAYLLTKDHIFLNSALRATAPYKFLESEQHGVKAVPNNKHDWYEEYPTTPSSF
                     VLNGFMYSILGIYLPLKEFAGEKLGKEARSLYERGMESLKAMPLYDTGSGTIYDLRHF
                     MLGIAPNLARWDYHTTHINOLOLLSTIDESPIFKEFVKRWKSYLKGSRAKHN"

BASE COUNT      903 a    625 c    654 g    903 t
ORIGIN

Query Match       100.0%;  Score 1404;  DB 7;  Length 3085;
Best Local Similarity  100.0%;
Matches 1404;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```
Qy    1  tccaagctgaattcctcatagctattccaaagtgtcaagcacagagagagccccttatcaccct   60
         ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  TCCAAGCTGAATTCTCATAGCTATTCCAAAGTGTCAAGCACAGAGAGAGCCCCTTATCACCCT   60

Qy   61  gatggtgtgtttatgtccttgaaggctacaaatgtggaagtccgagacgagtcaagtgc   120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  GATGGTGTGTTTATGTCCTTGAAGGCTACAAATGTGGAAGTCCGAGACGAGTCAAGTGC   120

Qy  121  ataagtggggttgaaggtgtaccttttatctcacacagtggggaccccaaggctattctac   180
         |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db  121  ATAAGTGGGGTTGAAGGTGTACCTTTTATCTCACACAGTGGGGACCTCAAGGCTATTCTAC   180

Qy  181  ccaatccagatctgcacagtatggttaagtcactacagcaagaaactccaagaatgaaaaccc   240
         |||||||||| ||||||||||||||||||||||||||||||||||| ||||  |||||||||||
Db  181  CCAATCCAGATTCTGCACAGTATGGTTAAGTCACTACAGCAAGAAATCTAACTGAAAAACCC   240

Qy  241  cctccatatagaggtatatgaaacagcaagaaacacagagaaacaagcaagcccaatgac   300
         |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  CCTCATATAGAGGTATATGAAACAGCAAGAAACACAGAGAAACAAGCAAGCCCAATGAC   300

Qy  301  tggactgtgcccaaggggctgctttattggctagtgtgtggctgatataagtcaagattcaccat   360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db  301  TGGACTGTGCCCAAGGGGCTGCTTTATTGGCTAGTGTGTGGCTGATATAAGTCAAGATTCACCAAT   360

Qy  361  gttaaacagttcattgctccaagaaaccagtgaaggtgtatcctttgcaactgggggaacaca   420
         |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||| |||||
Db  361  GTTAAACAGTTCATTGCTCCAGAAACCAGTGAAGGTGTATCCTTTGCAACTGGGGGACACA   420

Qy  421  aaagatttatttattttcattttgacctcaagttcttaacaaatggaagcgtgtctgtggtt   480
         |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  AAAGATTTTATTTATTTTCATTTTGACCTCAAGTTCTTAACAAATGGAAGCGTGTCTGTGGTT   480

Qy  481  ctggagacgacagaaaagaatcagctctcactgtacattatgtctcaaataacccagcta   540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db  481  CTGGAGACGACAGAAAAGAATCAGCTCTCACTGTACATTATGTCTCAAATACCCAGCTA   540

Qy  541  attgctctttaaagaagagacatatactatgtcatcgggcccagaacaatcatcggagcaca   600
         ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  ATTGCTTTTAAAGAAGAGACATATACTATGTCATCGGGCCCAGAACAATCATCGGAGCACA   600

Qy  601  gttaccccgggacctggtcactgacctcaggaaaggaagagttgggtctttccaacaccaaaagct   660
         |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  GTTACCCGGGACCTGGTCACTGACCTCAGGAAAGGAAGAGTTGGGTCTTTCCAACACCAAAAGCT   660

Qy  661  gtcaagccacaagaataatgcccaagaaggtggttaggttgattggcggaaaggaaaggc   720
         |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db  661  GTCAAGCCACAACAAGAATAATGCCCAAGAAGGTGGTTAGGTTGATTGGCGGAAAGGAAAGGC   720

Qy  721  ttccttgacaacattacccatctctaccacagcccaacaggcgcctcctctcgctgccagt   780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721  TTCCTTGACAACATTACCCATCTCTACCACAGCCCAACAGGCGCCTCCTCTCGCTGCCAGT   780

Qy  781  gactgggctggttgaggaaccacggggctggccgattatcggttgaccctgaccctgtaag   840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  781  GACTGGGCTGGTTGAGGAACCACGGGGCTGGCCGATTATCGGTTGACCCTGTAAG   840

Qy  841  ttaggggaaggcttcaagtctttagagccaggtggtactccggccaatggccaaggggcaa   900
```

```
RESULT  2

LOCUS       AK002188      2121 bp    mRNA            PRI      22-FEB-2000
DEFINITION  Homo sapiens cDNA FLJ11326 fis, clone PLACE1010481, highly similar
            to Homo sapiens mRNA for KIAA0836 protein.
ACCESSION   AK002188
VERSION     AK002188.1  GI:7023912
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE1
            clone:PLACE1010481.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
AUTHORS     Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,K.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
TITLE       NEDO human cDNA sequencing project
JOURNAL     Unpublished (2000)
REFERENCE   2  (bases 1 to 2121)
AUTHORS     Isogai,T. and Otsuki,T.
TITLE       Direct Submission
JOURNAL     Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1..2121
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE1010481"
                     /clone_lib="PLACE1"
                     /tissue_type="placenta"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      654 a    411 c    453 g    603 t
ORIGIN
```

Query Match      89.0%;  Score 1249.6;  DB 89;  Length 2121;
Best Local Similarity  94.1%;  Pred. No. 0;
Matches 1310;  Conservative  0;  Mismatches  79;  Indels  3;  Gaps  1;

```
Qy       13 tctcatagctattcccaaagtctatgcacagagagcccttacaccctgatggttgttt  72
            |||||||||||||||||||| |||||||| ||||||||||||||||| |||||||||||||
Db       52 TCTCATAGCTATTCCAAAGTCTATGCACAGAGAGCCCCTATCACCCCGATGGTGTGTTT  111

Qy       73 atgtcccttgaaggtcacaatgttgaagtcccgaagacagagagtgcaagtgcataagtggggtt  132
            ||||| |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db      112 ATGTCTTTGAAGGTCACAATGTGGAAGTCCGAGACAGAGTGCATAAGTGGGGTT  171

Qy      133 gaaggtgtgacctttactacacacagtggggaccctccaaggctatttctacccaatccagatt  192
            |||| |||||||||||| ||| ||||||||| ||||||||||||||||||||||||||||||||
Db      172 GAAGGTGTGACCATTATTACACACAATGGGGACCCTCACAGGCTATTTCTACCCAATCCAGATT  231

Qy      193 gcacagtatgggttaagtcactacagcaagaatctaactgaaaaacccctcatatagag  252
            ||||||||| ||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      232 GCACAGTATGGGATTAAGTCATTACACAGCAAGAATCTAACTGAGAAAACCTCTCACAATAGAG  291

Qy      253 gtatatgaaacagcagaagacagggggacaaaaacagcagcccaaatgactgactgtgccc  312
            |||| ||||||||||||||| ||||||||| ||||||| ||||||||||||||||||||||
Db      292 GTATATGAAACAGCAGAAGACAGGAGACAGACAAAAACACAAGAGACAGAAAAAA---CAAGCCTAATGACTGGACTGTGCCA  348

Qy      313 aagggctgcttatggctagtgtggctgaataagtcaagattcaaccaatgttaaaacagttc  372
            |||||||||||||||||||||||||||||| ||||||||||||||||||| |||||||||||
Db      349 AAGGGCTGCTTTATGGCGGAATGTGGCTGAATGGGCTGATAAGTCTCAGATTTCAGCACCAATGTCAAAACAGTTT  408

Qy      373 attggctcccagaacccaagtcagtgaaggtgtatcctcgaactgggaacacaaaaagatttatt  432
            |||||||||||||||||| ||||||| |||||||||||||||| ||||||||||||||||||| |
Db      409 ATTGCACCAGAAACCAGTGAAGGTGTATCCTTGCAACTGGGAACACAAAAAGATTTTATT  468

Qy      433 atttcatttgacctcaagttcttaacaaatggaagcgtgtctgttggttctcagccaaca  492
            |||||||||||| ||||||||||||| ||||| |||| |||||||||||||||||||||||
Db      469 ATTTCATTTGACCTCAAGTTCTTCACAAATACCAGTAGGAAGCGTGTCCGTGGGTTCTCAGACGACA  528

Qy      493 gaaagagaatcagcctctcactgtacattatctcaaataccagctaattgcttcttaa  552
            ||||||||||||||| |||||| ||||||||||||||||| || ||||||||||||||
Db      529 GAAAAGAGAATCAGCCTCTTTATGCGGAATGTGGCTGAATGGGCTGATAAGTCTCAGATTTCAGCACCAATGTCAAAACAGTTTTAAA  588

Qy      553 gaaaagagacatatactatggcatcggggcccagaacatcatggagcacagttacccgggac  612
            ||||||||||||| |||||||||||||||||||||||||||||||||| |||||||||||||
Db      589 GAAAAGAGACATATATACTATGGCATTGGCCCCAGAACTTCATGGAGCACAGTTACCCAGGGAC  648

Qy      613 ctggtcactgacctcaggagaaaggagtgggtctcttccaacacaaaaagctgtcaagccaaca  672
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db      649 CTGGTCACTGACCTCAGGAGAAAGGAGTGGGTCTTTCACAACACAAAAAGCTGTCAAGCCAACC  708

Qy      673 agaataatgcccaagaagtggttaggttggtttgattggaaggaggggcttcctgacac  732
            ||||| |||||||||||||||||||| |||||||||||||| |||||||||||||||||
Db      709 AGAATAATGCCCAAGAAGTGGTTAGGTTGGTTTGATTGGAAGGATTCCTGACAC  768

Qy      733 attacaatctctacacacagcccacatggctcgctctccgctcttcagctgactggctggtc  792
            |||||||||| ||||||||| ||||||| ||||||||| ||||||||||||||||||||||
Db      769 ATTACCAATCTCTACCACAGCCCACAATGGCTGCATTTTTTTGCCTGCTAGTGATTGGCTAGTA  828

Qy      793 aggaaccaggatgaggaaaggcggctggcgattatggtgacccgtaagttaggggaaggc  852
            |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db          
```

```
LOCUS       AB020643    4791 bp    mRNA       PRI       16-JUN-1999
DEFINITION  Homo sapiens mRNA for KIAA0836 protein, partial cds.
ACCESSION   AB020643
VERSION     AB020643.1   GI:4240160
KEYWORDS    .
SOURCE      Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
            SK plus clone:hj06165.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirosawa,M.,
            Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            XII. The complete sequences of 100 new cDNA clones from brain which
            code for large proteins in vitro
  JOURNAL   DNA Res. 5 (6), 355-364 (1998)
  MEDLINE   99156230
REFERENCE   2  (bases 1 to 4791)
  AUTHORS   Ohara,O., Suyama,M., Kikuno,R., Nagase,T. and Ishikawa,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-1998) to the DDBJ/EMBL/GenBank databases. Osamu
            Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3914,
            Fax:+81-438-52-3913,
FEATURES             Location/Qualifiers
     source          1..4791
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hj06165"
                     /clone_lib="pBluescriptII SK plus"
                     /dev_stage="adult"
                     /sex="male"
                     /tissue_type="brain"
     gene            1..1831
                     /gene="KIAA0836"
     CDS             <1..1831
                     /gene="KIAA0836"
                     /codon_start=2
                     /product="KIAA0836 protein"
                     /protein_id="BAA74859.1"
                     /db_xref="GI:4240161"
                     /translation="NYKTLIIICALFTLVTVLLMNKCSSDKALQFPRRSSSGFRVDGF
                     EKRAAASESNNYMNIVAKQQSEEAPPQEQOKAPPVVGGFNSNVGSKVLGLKYEEIDCL
                     INDHTIKGRREGNEVFLEPFTWVEKYFDVYGKVVQYDVDREFEHSYSKVVAQRAPY
                     HPDGVFMSFEGYNVEVRDRVKCISGVEGVPLSTOWGPQGYFYPIQIAQYGLSHYSKNL
                     TEKPHIEVVETAEDRDKNKPNDWYPKGCFMANVADKSRFTNVKQFTAPETSEGVSL
                     QLGNTKDFIISFDLKFLTNGSVSVVLETTENXLQLFTIHYVSNAQLIAFKERDIYYGIG
                     PRTSNSTVTRDLVIDLRKGVGLSNTKAVKPTKIMPKKVVRLIAKGKGFLDNIIISTTA
                     HMAAFFAASDMLVRNQDEKGGWPINVTRKLGEGFKSLEPGWYSAMAQQGAISTLVRAY
                     LLTKDHIFINSALRATAPYKFLSEOHGVKAVFMNKHDWYEEYPTTPSSPFVLNGPMYSL
                     IGLYDLKETAGEKLGKEARSLYERGMESLKAMLPLYDTGSGTIYDLRHFMLGIAPNLA
                     RWDYHYTHINQLLSTIDESPIFKEFVKRWKSYLKGSRAKHN"
BASE COUNT     888 a    950 c    1512 t
ORIGIN
           1441 a
```

```
Qy          3

RESULT   3

Db   1429 GCAAAGCACAC 1440
Qy   1393 gcaaagcacaac 1404
Db   1429 GCAAAGCACAC 1440
```

Query Match      89.0%;  Score 1249.6;  DB 85;  Length 4791;
Best Local Similarity   94.1%;
Matches 1310;  Conservative   0;  Mismatches   79;  Indels   3;  Gaps   1;

```
Qy   13 tctccatagctattccaaagtctatcgcacagagagagaccccttatcaccctgatgtgtgtt 72
        ||||| ||||||||||||||||| ||| |||||||||||||||||||||||||||||||||||
Db  440 TCTCATAGCTATTCCAAAGTCTATCGCACAGAGAGAGACCCCTTATCACCCTGATGGTGTGTTT 499

Qy   73 atgtcctttgaaggtcaaagtgtgggaagtccgagacagagagtcaaagtgcataagtggggtt 132
        ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  500 ATGTCTTTTGAAGGCTACAAATGTGGAAGTCCGAGACAGAGAGTCAAAGTGCATAAGTGGGGTT 559

Qy  133 gaaggtgtacctttattctacacagtgggggacctcaaggtctattcctacccaatccagatt 192
        ||||| |||| |||||||||||||||||||||||||||||||||||| |||| ||||||||
Db  560 GAAGGTGTGCCATTATCTTACACAATGGGGACCTCAAGGCTATTTCTATCCAATCCAGATT 619

Qy  193 gcacagtatgggttaagtcactacacagcaagaaatctaacctgaaaaacccccctcatatagag 252
        ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  620 GCACAGTATGGGTTAAGTCATTAAGTCATTACACAATGGGGCACCTCAAGAGAATCTAACTGAGAAAACCTCCTCACATAGAG 679

Qy  253 gtatatgaaacagcagaagacagggacaaaaacagcaagcccaatgactgactgtgctgccc 312
        ||||| |||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db  680 GTATATGAAACAGCAGAAGACAGAGAGACAGAACAAAA---CAAGCCTAATGACTGACTGTGCCA 736

Qy  373 attgctccagaaacctgaaaggttgatatcctgcaactgtggggacacacaaaagaagatttatt 432
        ||||||||||||||||| ||||||| ||||||||||||||||||||||||||||| ||||
Db  737 AAGGGCTCGCTTCTTATGGCCAAATGTGGGCCGATAAAGTCTAGAATTCAACAAGTCAAACAGTTT 796

Qy  433 atttcatttgacctcaagttcttaacaaaatggaagccgtgtctgtggttctggagacgaca 492
        ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  797 ATTTCATTTGACCTCAAGTTCTTAACAAAATGGAAGCCGTGTCTGTGGTTCTGGAGACCACA 856

Qy  493 gaaagaaatcagctctcaactgtacattatgtctcaaataccagctaattgctttaaa 552
        ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  857 GAAAGAAATCAGCTCTCAACTGTTACACTATACATTATGTCTCAACTATACATTATGTCTCAGCCTAATTGCTTTTAAA 916

Qy  553 gaaagagacatactatggcatcgggcccagaacacatggagcacagttacccgggac 612
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  917 GAAAGAGAATCAGCTCTTCACTATACATTATGTCTCAACTATACATTATGTCTCAGCCTAATTGCTTTTAAA 976
```

```
Qy  829 AGGAACCAGGATGAGAAAAGTGGTGACTCCGCAGGCCAATTATTGGTGCACCGTRAAGTTAGGGGAAGG 888
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  853 ttcaagtcttagaggcaaggggtggtactccgcaaggygcaagccattctaca 912
        ||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  889 TTCAAGTCTTTAGAGGCCAGGATGGTATTCTGCCATGGCCAAAGGCAAGGCCATTCTACA 948
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  913 ttagtcaggcctatctcttaaacaaaagaccatatattcctcaattcagctttaagggca 972
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  949 TTAGTCAGGCGCCTATCTGTTAACAAAAGACCATATATTCCTCAATTCAGCTTTAAGGGCA 1008
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  973 acagcccttacaagtttctgtcaagagcaacatgggagtcaaggctgttgttatgaataa 1032
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1009 ACAGCCCTTATTAAGTTTCTRATCTACAGCAGCAGTGAGTTGAGTTAAAGCTGTGTTTATGAATAAA 1068
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1033 catgagtggtatgaagaatatccaactacacaactagtcctcttttgtttcaaaatggctttatg 1092
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1069 CATGACTGGTATGAAGAATAATCCAACCACTAGCTCTTTTGTTTTCAAATGGCTTTATG 1128
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1093 tattcttaattgggtgtataggttgagcttaaagagaaactcgggagaaga 1152
        |||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1129 TATTCTTTAATTGGGCTGTATGGACTTAAAAGAAACTCGAGGGAAAAACTCGGAAAAGAA 1188
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1153 gcgaggtccttgtatgagctgcatggaatccctaaagccaatgctcccccttgtacgac 1212
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1189 GCAAGGTCCTTGTTGATGGCGTGGGCAAATCTTCTTAAAAGCCATGGTGTGCCCCTTTGTATGAC 1248
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1213 actggctcaaggaaccatctatgaccctccggcaacttcatgcttggcattgcccccaacctg 1272
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1249 ACTGGCGTCAGGAACCANCCATCTATGACCTCCGTCACTTCAGTGCTTGGCANCGCTCCTAACCTG 1308
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1273 gccgcgctgggagctatacacacaacaatcaactgcagctgcttagcacccattgat 1332
        ||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1309 GCTCGCGTCGGGACTATACAATCAATCAATCAGTTGCAGCTACTACGTACCCATTGAT 1368
        |||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1333 gagtcccaatcttcaaagaagttgtcaagggctacctcactgggcagccgg 1392
        |||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 1369 GAGTCCCCAATCTTCAAAGAAGAATTTCAAGAAGAATTGTGTCAAAGAGTGTGGGAAAAGTACCTTACGGAAACAAGGCCAGCAGG 1428
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
LOCUS       AF325532               2088 bp    mRNA            ROD       22-DEC-2000
DEFINITION  Mus musculus heparin/heparan sulfate:glucuronic acid C5 epimerase
            mRNA, complete cds.
ACCESSION   AF325532
VERSION     AF325532.1  GI:11935176
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2088)
  AUTHORS   Crawford,B.E., Olson,S.K., Pinhal,M.A.S. and Esko,J.D.
  TITLE     Cloning, Golgi Localization, and Enzyme Activity of the Full Length
            Heparin/Heparan Sulfate:Glucuronic Acid C5 Epimerase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2088)
  AUTHORS   Crawford,B.E., Olson,S.K., Pinhal,M.A.S. and Esko,J.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-2000) Cellular and Molecular Medicine, University
            of California, San Diego, 9500 Gilman Drive, La Jolla, CA
            92093-0687, USA
FEATURES             Location/Qualifiers
     source          1..2088
                     /organism="Mus musculus"
                     /strain="LAF1"
                     /db_xref="taxon:10090"
                     /tissue_type="Furth mastocytoma"
     CDS             232..2088
                     /note="heparan sulfate/heparin biosynthesis"
                     /codon_start=1
                     /product="heparin/heparan sulfate:glucuronic acid C5
                     epimerase"
                     /protein_id="AAG42004.1"
                     /db_xref="GI:11935177"
                     /translation="MRCLAARVNYKILIIICALFLTVLLMNKCSSDKAIQFPRHLS
                     SGFRVDGLEKRAASESNHYANHIAKQQSERAFPQEQQKAPPVVGGFNSNGGSKVLGL
                     KYEEIDCLLNDEHTIKGRECRENEVFLPFTWVEKYFDVYGKYVQVQDGYDREFFSHSYSK
                     VYAQRSPYHPDGVFMSFEGYNVEVRDRVKCISGVEGVPLSTQWGPQGYFPIQIAQYG
                     LSHYSKNLTEKPPHIEYETEADRDRNIRPNEWTVPKGCEMASVADKSRSTNVKQFIA
                     PETSEGVSLQLGNTKDFIISFDLLKLLTNGSVSVVLETTEKNQLFVHYHYVSNTQLLIAFR
                     DRDIYVGIGPRTSWSTVTRDLVTDLRKGVGLSNTRAVKPFKIMPKKVVRIAKGGFL
                     DNITISTTAHWAAFFAASDWLVRNQDEKGGNPIMVTRKLGEGFKSLEPGWYSAMAQGQ
                     AISTLVRAYLLLTKDYVFLSSALRAPAYPKFPSEQHGVKAYVEMNKHDWYEEYPTPSSF
                     VLNGFMYSLIGLYDLKETAGEFLGKEARPLYERGMESLKAMLPLYDTGSGTIYDLRHF
                     MLGIAPNLARDYHTHINQLLSTIDESPIFKEFVKRWKSYLKGSRAKHN"
BASE COUNT      601 a      470 c      503 g      514 t
ORIGIN
```

```
Qy     13 tctcaTAGctatttccaaagtctcaaagtcacagagagagagggagcaccctaccctgatggtgtt 72
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    694 TCTCATAGCTACTTTCCAAAGTCTATTCCAAAGTCTATGCACAGAGATCACCTTATCACCCTGACGGGTGTGTT 753

Qy     73 atgtccttgaaggtctacaaatgtggaaagtccgagacagagatcaaagtgcaaatgggggtt 132
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    754 ATGTCCTTGAAGGCTACAAATGTGGAAGTCCGAGACAGAGTCAAATGTATAAGTGGAGTT 813

Qy    133 gaaggtgtaccttatctcacacagctggggaacctcaaggctatttctacccaatccagatt 192
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    814 GAAGGTGTGCCATTAYCTACCCAGTGGGGCCCTCAAGGCTATTTCTACCCAATCCAGATT 873

Qy    193 gcacagtatgggttaagtcacacagcaagaaatcctaactgaaaaaccccctcatatagag 252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    874 GCACAGTATGGGCTAAGTCAYTACACAGCAAGAAATCTAACCGAGAAAAACCCCTCACATAGAA 933

Qy    253 gtatatgaaacgcagaagacaaaacagcaaagccccaatgactggactgtgtgcc 312
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    934 GTATATGAAACAGCAGAAGACAGGAAGACGAAACATCAGACCTAATGAATGGACTGTGCCC 993

Qy    313 aagggtgcttatatgctagtgtggctgataagtcaaagtcaccaaatgttaaacagttc 372
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    994 AAGGGGTGTTATATGGCTTCATGGCCAGTGGTGGCGACAAGTCTAGAACCACCAAATGTTAAACAGTTT 1053

Qy    373 attgtgtcccagaaaacaagtgaaggtgtatcctttgcaactgggaacacaaaagatttatt 432
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1054 ATTGCTTCCAGAAACACCAGTGAAGGTGTGTCTTTTGCAGCTGGGAACACAAAAGATTTCATT 1113

Qy    433 atttcaattgacctcaagttcttaacaaatgggagcgttgtctggttcctggagacgaca 492
          |||||||||||||||||||| |||||||||||||||||||||||||||||| |||
Db   1114 ATTTCAATTGACCTCAAGCTTTTAACAAATGGGAGCGTGTCTGTTGGTGTTCTGGAGACCACA 1173
```

```
     977 GAAAGAGAGATATATACTATGGCATTGGGCCCCAGAACTTCATGGGAGCACAGTTACCAGGGAC 1036
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     613 ctggtcactgacctcaggagaaagggaggtggtggtcttcccacacacaaagctgtcaagccaaca 672
    1037 CTGGTCACTGACCTCAGGAGAAAGGGAGGAGTGGGGTCTTCTTCAAACACACAAAGCTGTCAAGCCAACC 1096
     673 agaataaatgcccaagaagaagtgtgttaggttgattcgcaaagggaaggggcttcctttgacaac 732
    1097 AAAATAAATGCCCAAGAAGGTGGTTAGGTTGATTCGAATTCCTCGACAAC 1156
     733 attaccatctctaccacagcgcccacacaatggctgcttcctctcgctcgcctggctggtg 792
    1157 ATTACCATCTCTACCACACAGCCCACATGGCTGCTTCCTCTCGCTCGCCTGGCTACGTA 1216
     793 aggaaccaggagtgaagaaggggctggggccgatatatgggtggtgtggttaggggaaggc 852
    1217 AGGAACCAGGATGAGAAAAGGTGGCGGCCCGCAATTATGGGTGTAGTGTGGTTAGGGGAAGGG 1276
     853 ttcaagtctttagagccaggtggtactccgcgcatggcccaaatgatccccattttctaca 912
    1277 TTCAAGTCTTTAGAGCCAGGATGTGTATTCTGCCATGGCCCAAATGCCCAATTTCTACA 1336
     913 ttagtcaggggcctatcctctaacaaagaccaatatatctcaattcagctttaagggca 972
    1337 TTAGTCAGGGGCCTATCTGTGTTAACAAAAAGCAATATATCTCAATTCAGCTTTAAGGGCA 1396
     973 acagcccttacaagttcctgtcagagcagcagccaatgggactgtcaacccagtgggaggc 1032
    1397 ACAGCCCTTATAAGTTCTTATCTGAGCAGCAGCCAATGGGACTGTCAACCCAGTGGGAGGG 1456
    1033 catgactggttataagaagaatatccaactacacctactgctcctttgtttcaaatggctttatg 1092
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
    1457 CATGACTGGTATGAAGAGAATATCCAACCACACTTAGCTCTTTGTTTCAAATGGCTTTATG 1516
    1093 tattctttaattgggctgtatgtgacttaaagaaaactcgggaaataggga 1152
    1517 TATTCTTTAATTGGGCTGTATGACTTAAAGAAAAACTCGAGGAAAAACTCGGAAAGAA 1576
    1153 gcgaggtccttgtatgagcgtggcatggaatccccttaaaagccaatgcctgcccttgtacgac 1212
    1577 GCAAGGTCCTTGTATGAGCGTGGCATGGAATCTCTTAAAGCCATGCTTCCCTTGTATGAC 1636
    1213 actggtctcaggaaccatctatgtgactccgggccacttcatgtgccccaacctg 1272
    1637 ACTGGCTCAGGGACCATCTATGACCTCCGTCACTTCATGCTTGGCATCGCCCTGACGGGTGTGTT 1696
    1273 gcccgctgggactatcacacacacacacaactcaactcaactgcagctgctagcaccattgat 1332
    1697 GCTCCGCTGGGACTATCACACACACACACCATCAATCAGTTGCAGCTACTCAGTACCATTGAT 1756
    1333 gagtcccaatcttcaaaagaattgtcaagagaggtggaagagctacacttaaaggcagccgg 1392
    1757 GAGTCCCAAATCTTCAAAAGAATTTGTCAAGAAAGCTACTTAAGGCAGCAGG 1816
    1393 gcaaagcacaac 1404
         ||||||||||||
    1817 GCAAAGCACAAC 1828
```

```
Qy    493 gaaaggaatacaagctcttcaactgttacacattatgtctcaaatacccagctaattgctcttaaa 552
          |||| |||| |||||||||||| ||| |||||||||||||||| |||||||||| |||  ||| |
Db   1174 GAAAAAGAATCAGCTCTTCACTGTGCATTATGTCTCAAACACCCAGCTGATTGCTTTCAGA 1233

Qy    553 gaaagagaacatatactatggctcgggcgcccagaaacatcatgggacacagttaccccgggac 612
          ||  ||| ||| ||| ||||||| ||| ||||||||  |||||| |||| ||||| || ||| |
Db   1234 GACAGGGACAATANTACTACGGCATTNGGCCCCAGAACTTCATCGAGTTACCAGAGAC 1293

Qy    613 ctggtcaactgacctcaggaaaggagtggtcttccacacacaaaagctgtcaaagccaaca 672
          |  | ||  | |  |||||| |||| |||||| |||||  |||||| | ||||||||||||
Db   1294 AAAATCANTCATTGCCACCAANAAGGTGGTTAGGTTNGAATTGCAAAAGACTTGTCAAAGCCAACC 1353

Qy    673 agaatggctcccaaggaaggtggttaggttgttcggaaaggaaggcttccttgacaac 732
          |   |||  ||| |  ||||||||||||||| |  |  || ||| | ||| ||||||||
Db   1354 GCCAAAAAGGTGGTTAGGTTNGATTGCAAAAGGGAAGGGGGGATTCCTGGACAAC 1413

Qy    733 attacacatcttcaccacccacactggcttgcctctgtccagtgactggctggtg 792
          |||||  ||||| ||||| ||||| |||| |  | |  ||| |||| |||||| ||
Db   1414 ATTACCANTCTCAACCACAGCCCACATGGCTCGTTCATTCTTTGCTGCACAATGGCCTGGCTAGTG 1473

Qy    793 aggaaccaaggatgagaaaaggcggctcgccgattatcggttagttaggggaaggc 852
          |||||||| ||||||||||||| |||| || |  |||| ||  ||| |||| |||
Db   1474 AGGAACCAGGATGAGAAAAGGTGGCTGCCTGCCAATTATGGTGACCCGAAGTTACGGGAGGG 1533

Qy    853 ttcaagtctcttagagccaggagtggtactccgcccaaggcaagccaaccatttcaca 912
          ||   |||| ||| |||||||||||||||| | ||| ||| |||||||||| | ||||
Db   1534 TTTAAAATCTTTAGAACCAGGAGTGGTACTTCTGCCTGCCCAAGCCAAGCCAATCTCTACC 1593

Qy    913 ttagtcagggcctatcctcattaacaaaagaaccaatcccaattcaagctttaaggca 972
          |||||||||||||||| ||  || | ||| |  | ||  | ||  ||||||||| ||||
Db   1594 TTAGTCAGGGCCTATCTTCTAACGAAAAGACTATGTATTCCTCAGTTCAGCTTTAAGGCA 1653

Qy    973 acagcccttacaagttctctgtcagagcgaggttaaagccgtgttcatgaataaa 1032
          |||||||  ||||||| | || | ||||||| ||||||||||||||||| |||||
Db   1654 ACAGCCCCATACAAGTTTCCGTCGAGCGAGGTTAAAGCCGTGTTCATGAATAAA 1713

Qy   1033 catgactggttatgaagaataccaactacacctagctctcttgtttcaaatggctttatg 1092
          ||||||||| ||||| ||||||| | || |  |||||| | |||||| |||||||||||||
Db   1714 CATGACTGGTATGATGAAGAATATCAACCACTTAACAGAAGACCACCTAGCTCTCTTTTGTTTTAAATGGCTTTATG 1773

Qy   1093 tattccttaattgggctgtatgacttaaaagaaaacagacagggagacacttgggaaagaa 1152
          ||||| ||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   1774 TATTCTTTAATTGGGCTGTATGACTTAAAAGAAAACAGACAGGGAGACACTTGGGAAAGAA 1833

Qy   1153 gcgaggtccttgtatgagcgtggcaatggaatccctaaagcaatgccttaaagctgtacgac 1212
          || ||| |||||| |||||| | || ||||||| || |||||||||| | | | ||| | ||
Db   1834 GCAAGGCCCTTGTACGAGCGGCCGCCATGGAATCTCTTAAAGCATGTCCCTTGTTATCGAT 1893

Qy   1213 actggtcaggaaccacatctatggacctccggcacttcatgctgtcgcccccaaactg 1272
          ||||| | ||| ||| ||||| |||||||| |||||| ||| ||| ||| || ||||
Db   1894 ACTGGCTCCGGGACCANCATCTTATGACCTCCCCACTTCCATGGTGCTTCCCAACCTG 1953

Qy   1273 gcccgctgggactatcacccaccaccaatcatcaactgcagctgctagccaccattgat 1332
          ||||| ||||||||||||||||||||| | |   || ||| || | ||| |  |  |||
Db   1954 GCCCGGCTGGGACTATCACCCACCACCACCCAATTACCAGCTGCCCAGTTAACCTGCAT 2013

Qy   1333 gagtcccaatcttcaaagaatttgtcaaagaggtggaagctacccttaaaggcagccgg 1392
          |||||||||||||||| ||||||||||| ||| ||||| |||| ||||||||||| |||
Db   2014 GAGTCCCAATCTTCAAGAATTTGTCAAGAGTGGAAAAGCTACCTTTAAAGGCAGTAGG 2073

Qy   1393 gcaaagcacaac 1404
          ||||||||||||
Db   2074 GCAAAGCACAAC 2085
```

```
FEATURES             Location/Qualifiers
source               1..167187
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
misc_feature         1..1135
                     /clone_lib="RPCI-11 Human Male BAC"
                     /clone="RP11-352D13"
misc_feature         1236..2789
                     /note="assembly_fragment"
misc_feature         2890..4922
                     /note="assembly_fragment"
misc_feature         5023..6180
                     /note="assembly_fragment"
misc_feature         6281..7551
                     /note="assembly_fragment"
misc_feature         7652..9449
                     /note="assembly_fragment"
misc_feature         9550..11032
                     /note="assembly_fragment"
misc_feature         11133..12790
                     /note="assembly_fragment"
misc_feature         12891..15198
                     /note="assembly_fragment"
misc_feature         15299..17490
                     /note="assembly_fragment"
misc_feature         17591..20019
                     /note="assembly_fragment"
misc_feature         20120..23909
                     /note="assembly_fragment"
misc_feature         24010..27099
                     /note="assembly_fragment
                     vector_side:right"
                     clone_end:T7
misc_feature         27200..30782
                     /note="assembly_fragment
                     vector_side:right"
                     clone_end:SP6
misc_feature         30883..33932
                     /note="assembly_fragment"
misc_feature         34033..37691
                     /note="assembly_fragment"
misc_feature         37792..42856
                     /note="assembly_fragment"
misc_feature         42957..49339
                     /note="assembly_fragment"
misc_feature         49440..54849
                     /note="assembly_fragment"
misc_feature         54950..58121
                     /note="assembly_fragment"
misc_feature         58222..64249
                     /note="assembly_fragment"
misc_feature         64350..70780
                     /note="assembly_fragment"
misc_feature         70881..76628
                     /note="assembly_fragment"
misc_feature         76729..85215
                     /note="assembly_fragment"
misc_feature         85316..95800
                     /note="assembly_fragment"
misc_feature         95901..105851
                     /note="assembly_fragment"
misc_feature         105952..114860
                     /note="assembly_fragment"
misc_feature         114961..123755
                     /note="assembly_fragment"
misc_feature         123856..139927
                     /note="assembly_fragment"
```

```
*   6281  7551:  contig of 1271 bp in length
*   7552  7651:  gap of 100 bp
*         9449:  contig of 1798 bp in length
*   9450  9549:  gap of 100 bp
*   9550 11032:  contig of 1483 bp in length
*  11033 11132:  gap of 100 bp
*  11133 12790:  contig of 1658 bp in length
*  12791 12890:  gap of 100 bp
*  12891 15198:  contig of 2308 bp in length
*  15199 15298:  gap of 100 bp
*  15299 17490:  contig of 2192 bp in length
*  17491 17590:  gap of 100 bp
*  17591 20019:  contig of 2429 bp in length
*  20020 20119:  gap of 100 bp
*  20120 23909:  contig of 3790 bp in length
*  23910 24009:  gap of 100 bp
*  24010 27099:  contig of 3090 bp in length
*  27100 27199:  gap of 100 bp
*  27200 30782:  contig of 3583 bp in length
*  30783 30882:  gap of 100 bp
*  30883 33932:  contig of 3050 bp in length
*  33933 34032:  gap of 100 bp
*  34033 37691:  contig of 3659 bp in length
*  37692 37791:  gap of 100 bp
*  37792 42856:  contig of 5065 bp in length
*  42857 42956:  gap of 100 bp
*  42957 49339:  contig of 6383 bp in length
*  49340 49439:  gap of 100 bp
*  49440 54849:  contig of 5410 bp in length
*  54850 54949:  gap of 100 bp
*  54950 58121:  contig of 3172 bp in length
*  58122 58221:  gap of 100 bp
*  58222 64249:  contig of 6028 bp in length
*  64250 64349:  gap of 100 bp
*  64350 70780:  contig of 6431 bp in length
*  70781 70880:  gap of 100 bp
*  70881 76628:  contig of 5748 bp in length
*  76629 76728:  gap of 100 bp
*  76728 85215:  contig of 8487 bp in length
*  85216 85315:  gap of 100 bp
*  85316 95800:  contig of 10485 bp in length
*  95801 95900:  gap of 100 bp
*  95901 105851: contig of 9951 bp in length
* 105852 105951: gap of 100 bp
* 105952 114860: contig of 8909 bp in length
* 114861 114960: gap of 100 bp
* 114961 123755: contig of 8795 bp in length
* 123756 123855: gap of 100 bp
* 123856 139927: contig of 16072 bp in length
* 139928 140027: gap of 100 bp
* 140028 167187: contig of 27160 bp in length.
```

```
Query Match     66.5%;  Score 933;  DB 71;  Length 167187;
Best Local Similarity   93.3%;  Pred. No. 6.4e-245;
Matches  975;  Conservative  0;  Mismatches  70;  Indels  0;  Gaps  0;

Qy      360 tgttaaacagttcattgtctccagaaacccagtgaaggtgtatccttgcaactggggaacac 419
             |||| |||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   102759 TGTTGGTGTATATTCTCCCCTACAGAAACCCAGTGAAGGTGTATCCTTGCAACTGGGAAACAC 102818

Qy      420 aaaagatttttattttcatttgacctccaagttcttaacaaatggaaagcgtgtctgttggt 479
             |||||||||| || |||||||||||||||||||| |||||||||||||| ||||| |||||
Db   102819 AAAAGATTTTATTATTTCATTTGACCTCCAAGTTCTTGACAAATGGAAAGTGTGTCCGTGGT 102878

Qy      480 tctggagacgacagaaaagaaatccagctcttcactcgtacacattatgtctcaaataccccagct 539
             |||| ||||| || ||||| |||| ||||||||| ||||| |||||||| |||| ||||||||||
Db   102879 TCTAGAGACCACAGAAAAGAAATCACGTCCTTCACTATACARTAATGGTCTCARAATGTTCARGT 102938

Qy      540 aattcgtctttaaagaagagaacatatactatggatctcgggcccagaacaatcatcatgg 599
             |||| ||||||||||| ||||| |||||| |||| ||| ||||||||||||| |||| ||||
Db   102939 AATTGCTTTTAAAGAAAGAGATATATTACNTATNGGCACNTTGGGCCCAGAACTTCATTCATTGGAGAC 102998

Qy      600 agttacccgggacctggtcactgacctcaggagaaggagtcgggtctttccaacacacaaaagc 659
             ||||||| ||| |||||||||||||||||||| |||| ||| ||| ||||||| ||||||||||
Db   102999 AGTTACCAGGGGACCTGGTCACTGACCTCAGGAAAAGGGAGGAGTGGGTGGTCTTTCAAACACACAAAAGC 103058
```

```
Qy     660 tgtcaagccacaagaataatgccaagagaagtgttagttgattgccaaggaagg     719
           ||||||||||  |||||||||||| ||||||| ||| ||||| |||||||| ||  |||||
Db  103059 TGTCAAGCCAACCAAAATAATGCCCAAGAGAAGGTGGTTAGGTTGATTGCAAAAGGTAAGG  103118

Qy     720 cttccttgacacaattaccatctcctacacacagagaagaagagtgttagttggttcttcgctgcaag  779
           |||| | ||||| ||||||||||||||||||| ||  ||| |||| ||  ||| |||||||||  | |
Db  103119 ATTCCTCGACAACATTACCATCTCCTACACAGAGCCACATGGCCTGCATTTTTGCTGCTAG        103178

Qy     780 tgactggctggtgaggaaccagagaaggagaagggctggcccgattatggtgaccggtaa  839
           ||| |||||| |||||||||||| | ||||| |||| ||||||| |||| | |||||||||
Db  103179 TGATTGGCTAGTAAGGAACCAGGATGAGAAAGGTGGCTGGCCAATTAGTGTGACCGGTAA   103238

Qy     840 gttagggaaggcttcaagtcttagagcccaggtgtgttactccgcccatggccccaagggca  899
           |||||||||||| |||||  |||||||||||| || ||| | |||| ||||||||||||| ||
Db  103239 GTTAGGGAAGGGTTCAACTCTTTAGAGCCCAGGATGGTTATTCTCCACGGCCCCAAGGGCCA   103298

Qy     900 agccattctacattagtcaggggcctatcctcttaacaaagacccatatatattccaatcc  959
           ||||| |||||| ||||||||| |||||| | |||||||| |||||||||||||||| ||||
Db  103299 AGCCANTTCTACANTTAGTCAGGGCCTATCGTGTTAACAAAAGACCCATATATATTCCTCAATTC  103358

Qy     960 agcttaagggcaacagccccctacaagttctctgcaggacgaatggagtcaaggctgt     1019
           |||||||||| ||||||||| ||| ||||| |  |   ||| ||||||| | ||||||
Db  103359 AGCTTAAGGCAACAGCCCCTTATAAGTTTCTATCTGACCAGGATGGAGTTAAAGCTGT        103418

Qy    1020 gttatatgaataaacatgactggtgtatgaagaatatccaactacccagctgtctctttttt  1079
           ||| | ||||||||||||||||| | |||||||||| ||||| |||| |||  ||| || ||
Db  103419 GTTTATGAATAAAACATGACTGGGTAATGAAGAATAATCCAACCACACTACGTCTTTTGTTTTT    103478

Qy    1080 aaatggctttatgtattctttaattgggcgtatgacttagacttaaagaaactgcaggggaaa  1139
           ||||||||||||||| ||||||||||| | |||| |||||| |||||||||||||||||||||
Db  103479 AAATGGCTTTATGTAATTCTTTAATTGGCCTGTGTATGAGACTTAAAAGAAAACTGCAGGGGAAA  103538

Qy    1140 actcgggaaagacgaggagtgtcctttgtatagacgtggcatggaattcccttaaaagccatgct  1199
           |||| ||||| ||||| ||| ||  | |||| |||||||||||||||| |  |||||| |||||
Db  103539 ACTCCGGAAAAGACGAGAAGTTGTATGACGTGGCATGGAATCTCTTAAAGCCCATGCT          103598

Qy    1200 cccctgttacgacaactggctgcaggaacaatctatgaccctccgcgcacttcaatgcat  1259
           ||| ||||| |||| |||||| ||||||||||||||||||||||   | ||  | |||||
Db  103599 GCCCTGTTATGACACTGGCTTCAGGAACAATCTATGACCCTCCACTTGTCTTGGCAT      103658

Qy    1260 tgccccaacctggccgcgtggggactatatccacccaccaatcaatcaactgcagctgct  1319
           |||| | ||||| ||  |||||||| | ||||||||||| |||||| || | ||| | ||
Db  103659 CGCTCCTTACCTGCCTGGGACTATCATACCACCACCAATCAATCAGTGTTGCAGCTACT   103718

Qy    1320 tagcaccattgatgagtccccaatctctcaaagaaattgtcaaagaggtgaagaagactacct  1379
           ||| ||| |||||||||||||| || | |||||||| ||||||| |||||| || ||||||||
Db  103719 CAGTACCAATGATGAGTCCCCAGTCTTCAAAGAATTTGTCAAGAGAGGTGGAAAAAGCTACCT    103778

Qy    1380 taaaggcagcgcggcaaagcacaac  1404
           ||||||||| | |||||||||| |||
Db  103779 TAAAGGCAGAGGGCAAAAGCAAC    103803
```

RESULT  6

```
LOCUS       AC073380   171073 bp    DNA            HTG      28-JUN-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-82G5 map 17, WORKING DRAFT
            SEQUENCE, 22 unordered pieces.
ACCESSION   AC073380
VERSION     AC073380.2  GI:8780925
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171073)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-82G5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171073)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 28, 2000 this sequence version replaced gi:8567836.
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ------------------- Project Information
            Center project name: 82_G_5
            Center clone name: L10459
            ------------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 158402 bases at least Q40
            Consensus quality: 165231 bases at least Q30
            Consensus quality: 167603 bases at least Q20
            Insert size: 168973; sum-of-contigs
            Quality coverage: 4.0 in Q20 bases; sum-of-contigs
            ---------------------
Direct Submission
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1    1359: contig of 1359 bp in length
            *       1360    1459: gap of 100 bp
            *       1460    2848: contig of 1389 bp in length
            *       2849    2948: gap of 100 bp
            *       2949    3998: contig of 1050 bp in length
            *       3999    4098: gap of 100 bp
            *       4099    6461: contig of 2363 bp in length
            *       6462    6561: gap of 100 bp
            *       6562    8310: contig of 1749 bp in length
            *       8311    8410: gap of 100 bp
            *       8411   10590: contig of 2180 bp in length
            *      10591   10690: gap of 100 bp
            *      10691   12463: contig of 1773 bp in length
            *      12464   12563: gap of 100 bp
            *      12564   14454: contig of 1891 bp in length
            *      14455   14554: gap of 100 bp
            *      14555   16738: contig of 2184 bp in length
            *      16739   16838: gap of 100 bp
            *      16839   19574: contig of 2736 bp in length
            *      19575   19674: gap of 100 bp
            *      19675   21309: contig of 1635 bp in length
            *      21310   21409: gap of 100 bp
```

```
FEATURES             Location/Qualifiers
source               1..171073
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone_lib="RPCI-11 Human Male BAC"
misc_feature         1..1359
                     /note="assembly_fragment"
misc_feature         1460..2848
                     /note="assembly_fragment"
misc_feature         2949..3998
                     /note="assembly_fragment"
misc_feature         4099..6461
                     /note="assembly_fragment"
misc_feature         6562..8310
                     /note="assembly_fragment"
misc_feature         8411..10590
                     /note="assembly_fragment"
misc_feature         10691..12463
                     /note="assembly_fragment"
misc_feature         12564..14454
                     /note="assembly_fragment"
misc_feature         14555..16738
                     /note="assembly_fragment"
misc_feature         16839..19574
                     /note="assembly_fragment"
misc_feature         19675..21309
                     /note="assembly_fragment"
misc_feature         21410..27209
                     /clone_end:SP6
                     vector_side:right"
misc_feature         27310..34663
                     /note="assembly_fragment"
misc_feature         34764..41925
                     /note="assembly_fragment"
misc_feature         42026..48610
                     /note="assembly_fragment"
misc_feature         48711..58240
                     /note="assembly_fragment"
misc_feature         58341..69915
                     /note="assembly_fragment"
misc_feature         70016..83336
                     /note="assembly_fragment"
misc_feature         83437..96601
                     /note="assembly_fragment"
misc_feature         96702..113676
                     /note="assembly_fragment"
misc_feature         113777..136870
                     /note="assembly_fragment"
misc_feature         136971..171073
                     /note="assembly_fragment"
```

```
21410  27209: contig of 5800 bp in length.
27210  27309: gap of 100 bp
27310  34663: contig of 7354 bp in length
34664  34763: gap of 100 bp
34764  41925: contig of 7162 bp in length
41926  42025: gap of 100 bp
42026  48610: contig of 6585 bp in length
48611  48710: gap of 100 bp
48711  58240: contig of 9530 bp in length
58241  58340: gap of 100 bp
58341  69915: contig of 11575 bp in length
69916  70015: gap of 100 bp
70016  83336: contig of 13321 bp in length
83337  83436: gap of 100 bp
83437  96601: contig of 13165 bp in length
96602  96701: gap of 100 bp
96702  113676: contig of 16975 bp in length
113677 113776: gap of 100 bp
113777 136870: contig of 23094 bp in length
136871 136970: gap of 100 bp
136971 171073: contig of 34103 bp in length
```

```
misc_feature   136971..171073
               /note="assembly_fragment"
BASE COUNT     55739 a   30189 c   29971 g   53068 t   2106 others
ORIGIN
```

```
Query Match     34.9%;  Score 490.6;  DB 75;  Length 171073;
Best Local Similarity  94.6%;  Pred. No. 1.6e-123;
Matches  508;  Conservative  0;  Mismatches  29;  Indels  0;  Gaps  0;

Qy    868  ccagggtggtactccgccatggcccaagggcaagccaattctacattagtcagggcctat   927
           ||||||| |||| |||||||||||||||| ||||| |||||||||||||||||||||||||
Db   4099  CCAGGATGGTATTCTGCCATGGCCCAAGGGCAAGCCAATTCTACATTAGTCAGGGCCTAT  4158

Qy    928  ctcttaacaaaagaccacatatattcctcaattcagcttaagggcaagcccctacaag    987
           |||||||||||||||||||| || ||||||||||||||| ||||||||||||  ||| ||
Db   4159  CTGTTAACAAAAGACCATAATATTCCTCAATTCAGCTTTAAGGGCAACAGCCCCTTATAAG  4218

Qy    988  tttctgtcaggcagcatgggtcaaggctgttttatgaataaacatgactggtatgaa   1047
           |||| ||| |||||||||| ||| || |||| ||||||||||||||||||||||||||||
Db   4219  TTTCTATCTGAGCAGCATGGAGTAAAGCTGTGTTGTTATGAATAAACATGACTGGTATGAA  4278

Qy   1048  gaatatccaactacacctagtctctttgtttaaatggctttatgtatttcttttaattggg  1107
           |||||||||||| ||||||| |||| |||||||||||||||||||||||||||||||||||
Db   4279  GAATATCCAACCACACCTAGCTCTTTTGTTTAAATGGCTTTATGTATTTCTTTTAATTGGG  4338

Qy   1108  ctgtatgacttaaaagaaaactgcaggggaaaaactcgggaagaagcgaggtcctgtat   1167
           ||||||||||||||||||||| ||||||| |||||||||||||||||| |||||| ||||
Db   4339  CTGTATGACTTAAAAGAAAACTGCAGGGAAAAAACTCGGGAAGAAGCAAGGTCCTTGTAT  4398

Qy   1168  gagcgtggcatggaatcccctaaaagccatgctccccttgtacgacaactggctcaggaacc  1227
           |||||||||||||||||| || ||||||||| |||||||||||||| |||||||||||||
Db   4399  GAGCGTGGCATGGAATCTCTTAAAGCCATGCTGCCCCTTGTACGACACTGGCTCAGGAACC  4458

Qy   1228  atctatgaacctccggcacttcatgcttcttggcattgccccccaacctcgcccgctgggacttat  1287
           |||||||||||||| |||||||||||||||||||| |  |||  || |||||| |||||||||
Db   4459  ATCTATGAACCTCCGTCACTTCATGCTTCTTGGCATCGGTTCTTCCTTCCTCCCTGGGACTAT  4518

Qy   1288  cacacccacacaatcaatcaactgcagctgcttagcaccattgatgagtccccaatcttc  1347
           ||| ||||| |  ||||||||||| ||||| || |  ||||||||||||||| | ||||||
Db   4519  CATACCCACCCACCCAATCAATCACGTTGCAGCTACTCAGTTACCATTGATGAGTCCCACTCTTC  4578

Qy   1348  aaagaatttgtcaagagggtggaagagctacccttaaaggcacgccgggcaaagcacaac  1404
           |||||| |||||||||||||||||| |||||| |||||||||  | |||||||||||||||
Db   4579  AAAGAAATTTGTCAAGAGGGTGGAAAAGCTACCTTAAAGGCAGCAGGGCAAAGCACAAC  4635
```

```
RESULT  7
LOCUS       AC026992/c
ACCESSION   AC026992
DEFINITION  Homo sapiens chromosome 15 clone RP11-352D13

LOCUS       AC026992   167187 bp   DNA          HTG   03-JUL-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-352D13 map 15, WORKING DRAFT
            SEQUENCE, 30 unordered pieces.
ACCESSION   AC026992
VERSION     AC026992.4  GI:8901256
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167187)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-352D13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167187)
```
```
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,M., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Dearellano,I., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
```

```
  TITLE
  JOURNAL   Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,T.M, Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT   On Jul 3, 2000 this sequence version replaced gi:8079836.
            All repeats were identified using RepeatMasker.
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            ------------- Project Information
              Center project name: L8712
              Center clone name: 352_D_13
            ------------- Summary Statistics
              Sequencing Vector: M13; M77815; 100% of reads
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.960731
              Assembly quality: 147074 bases at least Q40
              Consensus quality: 15353 bases at least Q30
              Consensus quality: 15298 bases at least Q20
              Consensus quality: 161662 bases at least Q20
              Insert size: 163000; agarose-fp
              Insert size: 164287; sum-of-contigs
              Quality coverage: 3.5 in Q20 bases; agarose-fp
              Quality coverage: 3.4 in Q20 bases; sum-of-contigs
            --------------
            NOTE: This is a 'working draft' sequence. It currently
            * consists of 30 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *           1    1135: contig of 1135 bp in length
            *        1136    1235: gap of   100 bp
            *        1236    2789: contig of 1554 bp in length
            *        2790    2889: gap of   100 bp
            *        2890    4922: contig of 2033 bp in length
            *        4923    5022: gap of   100 bp
            *        5023    6180: contig of 1158 bp in length
            *        6181    6280: gap of   100 bp
            *        6281    7551: contig of 1271 bp in length
            *        7552    7651: gap of   100 bp
            *        7652    9449: contig of 1798 bp in length
            *        9450    9549: gap of   100 bp
            *        9550   11032: contig of 1483 bp in length
            *       11033   11132: gap of   100 bp
            *       11133   12790: contig of 1658 bp in length
            *       12791   12890: gap of   100 bp
            *       12891   15198: contig of 2308 bp in length
            *       15199   15298: gap of   100 bp
            *       15299   17490: contig of 2192 bp in length
            *       17491   17590: gap of   100 bp
            *       17591   20019: contig of 2429 bp in length
            *       20020   20119: gap of   100 bp
            *       20120   23909: contig of 3790 bp in length
            *       23910   24009: gap of   100 bp
            *       24010   27099: contig of 3090 bp in length
            *       27100   27199: gap of   100 bp
```

```
FEATURES             Location/Qualifiers
     source          1..167187
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-352013"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1135
                     /note="assembly_fragment"
     misc_feature    1236..2789
                     /note="assembly_fragment"
     misc_feature    2890..4922
                     /note="assembly_fragment"
     misc_feature    5023..6180
                     /note="assembly_fragment"
     misc_feature    6281..7551
                     /note="assembly_fragment"
     misc_feature    7652..9449
                     /note="assembly_fragment"
     misc_feature    9550..11032
                     /note="assembly_fragment"
     misc_feature    11133..12790
                     /note="assembly_fragment"
     misc_feature    12891..15198
                     /note="assembly_fragment"
     misc_feature    15299..17490
                     /note="assembly_fragment"
     misc_feature    17591..20019
                     /note="assembly_fragment"
     misc_feature    20120..23909
                     /note="assembly_fragment"
     misc_feature    24010..27099
                     /note="assembly_fragment"
     misc_feature    vector_side:right"
                     clone_end:T7
                     /note="assembly_fragment
     misc_feature    27200..30782
                     /note="assembly_fragment"
     misc_feature    30883..33932
                     /note="assembly_fragment"
```

```
    *       27200   30782: contig of 3583 bp in length
    *       30783   30882: gap of   100 bp
    *       30883   33932: contig of 3050 bp in length
    *       33933   34032: gap of   100 bp
    *       34033   37691: contig of 3659 bp in length
    *       37692   37791: gap of   100 bp
    *       37792   42856: contig of 5065 bp in length
    *       42857   42956: gap of   100 bp
    *       42957   49339: contig of 6383 bp in length
    *       49340   49439: gap of   100 bp
    *       49440   54849: contig of 5410 bp in length
    *       54850   54949: gap of   100 bp
    *       54950   58121: contig of 3172 bp in length
    *       58122   58221: gap of   100 bp
    *       58222   64249: contig of 6028 bp in length
    *       64250   64349: gap of   100 bp
    *       64350   70780: contig of 6431 bp in length
    *       70781   70880: gap of   100 bp
    *       70881   76628: contig of 5748 bp in length
    *       76629   76728: gap of   100 bp
    *       76729   85215: contig of 8487 bp in length
    *       85216   85315: gap of   100 bp
    *       85316   95800: contig of 10485 bp in length
    *       95801   95900: gap of   100 bp
    *       95901  105851: contig of 9951 bp in length
    *      105852  105951: gap of   100 bp
    *      105952  114860: contig of 8909 bp in length
    *      114861  114960: gap of   100 bp
    *      114961  123755: contig of 8795 bp in length
    *      123756  123855: gap of   100 bp
    *      123856  139927: contig of 16072 bp in length
    *      139928  140027: gap of   100 bp
    *      140028  167187: contig of 27160 bp in length.
```

```
misc_feature     34033..37691
                 /note="assembly_fragment"
misc_feature     37792..42856
                 /note="assembly_fragment"
misc_feature     42957..49339
                 /note="assembly_fragment"
misc_feature     49440..54849
                 /note="assembly_fragment"
misc_feature     54950..58121
                 /note="assembly_fragment
                 vector_side:right"
                 /note="assembly_fragment
                 clone_end:SP6"
misc_feature     58222..64249
                 /note="assembly_fragment"
misc_feature     64350..70780
                 /note="assembly_fragment"
misc_feature     70881..76628
                 /note="assembly_fragment"
misc_feature     76729..85215
                 /note="assembly_fragment"
misc_feature     85316..95800
                 /note="assembly_fragment"
misc_feature     95901..105851
                 /note="assembly_fragment"
misc_feature     105952..114860
                 /note="assembly_fragment"
misc_feature     114961..127755
                 /note="assembly_fragment"
misc_feature     123856..139927
                 /note="assembly_fragment"

Query Match    14.5%;  Score 204;  DB 71;  Length 167187;
Best Local Similarity  90.9%;  Pred. No. 6.8e-45;
Matches  229;  Conservative  0;  Mismatches  20;  Indels  3;  Gaps  1;

Qy    131  ttgaagtgtaccttttatctacacaacagtggggggacctgaaggcgctattcacaatccaga  190
           ||||||  ||||||||||||||||||||| |||||||||||| |||||||  ||||||||||
Db   6578  TTATAGGTGTGCCATTATCTACACAAATGGGGACCTCAAGGCTATTTCTATCCAATCCAGA  6519

Qy    191  ttgcacagtatgggttaagtcacacagcgaagaatctaactgaaaaacccctccatatag  250
           |||||||||||| |||||||||| |||||||||||||||||  |||||| ||| |||||
Db   6518  TTGCACAGTATGGATTAAGTCACTTACACAGCGAAGAATCTAACTGAGAAAACCTCTCACATAG  6459

Qy    251  aggtatatgaaacagcagggaagacaaacagcaagcccaatgactgactgtgc  310
           ||||||||||||||||| |  |||| ||||||||||||||||||||||||||
Db   6458  AGGTATATGAAACAGCAGCAGAGAGAGACAAAAA---CAAGCCTAATGACTGACTGTGC  6402

Qy    311  ccaagggctgcttatggctagtgtggctgataagtcaagattcaccaaatgttaaacagt  370
           ||||||||||| ||||||| ||||||||||||||| || |||||||||||| ||||||||
Db   6401  CAAAGGGCTGCTTTATGGCGGAATGTGGCTGATAAGTCTAGATTCACCAAATGTCAAACAGT  6342

Qy    371  tcattgtccag  382
           |||||| ||||
Db   6341  TTATTGCACCAG  6330

RESULT  8
ACO18117/c
LOCUS       ACO18117    29372 bp    DNA    HTG    09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   ACO18117
VERSION     ACO18117.1  GI:6553074
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 29372)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
```

```
COMMENT     Rockville, MD, USA
            This sequence was identified as CDM:10213417 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE:This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..29372
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     7866 a    6483 c    6794 g    8229 t
ORIGIN

Query Match    10.7%;  Score 150.8;  DB 65;  Length 29372;
Best Local Similarity  63.0%;  Pred. No. 2.6e-30;
Matches  233;  Conservative  0;  Mismatches  137;  Indels  0;  Gaps  0;

Qy   1035  tgactggttatgaagaatatcacaccacactagctctttttgttttaaatggctttatgta  1094
           ||  ||| || || |||| | | ||||||||| |  |  ||||| ||  || |||||| | ||
Db   5572  TTACAGGTTACAGGAGAGTACCCCACACCACCAGCCACCCTCTTATGTTGCTAAACGGCTTTATCTA  5513

Qy   1095  ttctttaattgggctgtgtatgacttaaagagaaactgcagggggaaaaactcggggaaagaagc  1154
           || | | |||| |||||| ||||| | ||| |  | |||||| ||||||| | | |||  |||
Db   5512  TTCACTGCTGGGGTCTCTACGAGTCTCAACGACACTGCCCCGGGAAAAATCGCCCGCGAGGC  5453

Qy   1155  gaggtccttgtatgagcgtggcattgaaatcccttaaagccatgctgctccccttgtacgacac  1214
           | || |  ||| | | | || ||| |  ||ccctaaa  |||||| | ||| || | ||| ||
Db   5452  TGGAAAGCTTTTCCCGCACAGGGCATGCACTCCCTTAAAAAGATGCTGTTGCTAATTGACAC  5393

Qy   1215  tggctcaggaaccatctatgacctccggcacttcatgcttggcaattgccccccaacctggc  1274
           |||||| || || |||| |||||| || || | ||| |  |||   || | |||  || |||
Db   5392  TGGCTCCGGCCGCACCAGCTACGACCTGCCCATCGCCTTCGAGCCCTGGGCGCTAGCTCCTCAAATCTGGC  5333

Qy   1275  ccgctgggactatcacacaccacaccacaatcaatcaactgagctgcttagcacacattgatga  1334
           ||||||||||| || || |||| | |||| |   || |||| || || |||||| |||| || |
Db   5332  GCCGCTGGGACTACCATGCACACACCATGCGAACAGGTGAAACAGTTACTTCGTTAGCCACCATCGACAG  5273

Qy   1335  gtccccaatcttcaaaagaaatttgtcaaggaggtgggaaggcgtacctaaaagcagcccggc  1394
           ||| ||||  || ||| |||||| ||||| |||| ||||| | | || ||||||||||| |||
Db   5272  CGATCCATTAATTGCACAAACTCGGAGCGCTGGAAGGGCGCTATATGTTTGGGTCGTGCTGCGGGC  5213

Qy   1395  aaaagcacaac  1404
           ||| |||||||
Db   5212  CAAGCACAAC  5203

RESULT  9
AC007593/c
LOCUS       AC007593    119597 bp    DNA    HTG    21-FEB-2000
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR01C10 (D620) RPCI-98
            01.C.10 map 42E-43A strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 89 unordered pieces.
ACCESSION   AC007593
VERSION     AC007593.4  GI:7018749
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 119597)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Sviskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 119597)
AUTHORS     Celniker,S.E., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,M., Hummasti,S., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
TITLE       Direct Submission
JOURNAL     Submitted (20-MAY-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 21, 2000 this sequence version replaced gi:4885670.
          * For further information about this sequence, including its location
          * and relationship to other sequences, please visit our sequence
          * archive Web site (http://www.fruitfly.org/sequence/) or send email
          * to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
          * the following cutoffs: length >= 200 bases.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 89 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.

          *        1      584: contig of 584 bp in length
          *      585      664: gap of unknown length
          *      665     1363: contig of 699 bp in length
          *     1364     1443: gap of unknown length
          *     1444     2181: contig of 738 bp in length
          *     2182     2261: gap of unknown length
          *     2262     3104: contig of 843 bp in length
          *     3105     3184: gap of unknown length
          *     3185     3906: contig of 722 bp in length
          *     3907     3986: gap of unknown length
          *     3987     4920: contig of 934 bp in length
          *     4921     5000: gap of unknown length
          *     5001     5966: contig of 966 bp in length
          *     5967     6046: gap of unknown length
          *     6047     6862: contig of 816 bp in length
          *     6863     6942: gap of unknown length
          *     6943     7461: contig of 519 bp in length
          *     7462     7541: gap of unknown length
          *     7542     8246: contig of 705 bp in length
          *     8247     8326: gap of unknown length
          *     8327     9098: contig of 772 bp in length
          *     9099     9178: gap of unknown length
          *     9179     9852: contig of 674 bp in length.
          *     9853     9932: gap of unknown length
          *     9933    10800: contig of 868 bp in length
          *    10801    10880: gap of unknown length
          *    10881    12098: contig of 1218 bp in length
          *    12099    12178: gap of unknown length
          *    12179    13330: contig of 1152 bp in length
          *    13331    13410: gap of unknown length
          *    13411    14082: contig of 672 bp in length
          *    14083    14162: gap of unknown length
          *    14163    14710: contig of 548 bp in length
          *    14711    14790: gap of unknown length
          *    14791    15379: contig of 589 bp in length
          *    15380    15459: gap of unknown length
          *    15460    16354: contig of 895 bp in length
          *    16355    16434: gap of unknown length
          *    16435    17123: contig of 689 bp in length
          *    17124    17203: gap of unknown length
          *    17204    18522: contig of 1319 bp in length
          *    18523    18602: gap of unknown length
          *    18603    19730: contig of 1128 bp in length
          *    19731    19810: gap of unknown length
          *    19811    20869: contig of 1059 bp in length
          *    20870    20949: gap of unknown length

          *    20950    22525: contig of 1576 bp in length
          *    22526    22605: gap of unknown length
          *    22606    23882: contig of 1277 bp in length
          *    23883    23962: gap of unknown length
          *    23963    25201: contig of 1239 bp in length
          *    25202    25281: gap of unknown length
          *    25282    26311: contig of 1030 bp in length
          *    26312    26391: gap of unknown length
          *    26392    27304: contig of 913 bp in length
          *    27305    27384: gap of unknown length
          *    27385    28957: contig of 1573 bp in length
          *    28958    29037: gap of unknown length
          *    29038    29951: contig of 914 bp in length
          *    29952    30031: gap of unknown length
          *    30032    30952: contig of 921 bp in length
          *    30953    31032: gap of unknown length
          *    31033    31672: contig of 640 bp in length
          *    31673    31752: gap of unknown length
          *    31753    32764: contig of 1012 bp in length
          *    32765    32844: gap of unknown length
          *    32845    34084: contig of 1240 bp in length
          *    34085    34164: gap of unknown length
          *    34165    35006: contig of 842 bp in length
          *    35007    35086: gap of unknown length
          *    35087    36296: contig of 1210 bp in length
          *    36297    36376: gap of unknown length
          *    36377    37977: contig of 1601 bp in length
          *    37978    38057: gap of unknown length
          *    38058    39491: contig of 1434 bp in length
          *    39492    39571: gap of unknown length
          *    39572    40645: contig of 1074 bp in length
          *    40646    40725: gap of unknown length
          *    40726    42319: contig of 1594 bp in length
          *    42320    42399: gap of unknown length
          *    42400    43547: contig of 1148 bp in length
          *    43548    43627: gap of unknown length
          *    43628    45038: contig of 1411 bp in length
          *    45039    45118: gap of unknown length
          *    45119    46182: contig of 1064 bp in length
          *    46183    46262: gap of unknown length
          *    46263    48155: contig of 1893 bp in length
          *    48156    48235: gap of unknown length
          *    48236    49922: contig of 1687 bp in length
          *    49923    50002: gap of unknown length
          *    50003    51972: contig of 1970 bp in length
          *    51973    52052: gap of unknown length
          *    52053    54342: contig of 2290 bp in length
          *    54343    54422: gap of unknown length
          *    54423    55950: contig of 1528 bp in length
          *    55951    56030: gap of unknown length
          *    56031    57106: contig of 1076 bp in length
          *    57107    57186: gap of unknown length
          *    57187    58601: contig of 1415 bp in length
          *    58602    58681: gap of unknown length
          *    58682    59984: contig of 1303 bp in length
          *    59985    60064: gap of unknown length
          *    60065    61596: contig of 1532 bp in length
          *    61597    61676: gap of unknown length
          *    61677    63106: contig of 1430 bp in length
          *    63107    63186: gap of unknown length
          *    63187    65066: contig of 1880 bp in length
          *    65067    65146: gap of unknown length
          *    65147    67491: contig of 2345 bp in length
          *    67492    67571: gap of unknown length
          *    67572    71138: contig of 3567 bp in length
          *    71139    71218: gap of unknown length
          *    71219    73245: contig of 2027 bp in length
          *    73246    73325: gap of unknown length
          *    73326    76254: contig of 2929 bp in length
          *    76255    76334: gap of unknown length
          *    76335    79184: contig of 2850 bp in length
          *    79185    79264: gap of unknown length
          *    79265    83109: contig of 3845 bp in length
```

```
RESULT  10
AC007624/c

LOCUS       AC007624   167062 bp   DNA         HTG       13-DEC-1999
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR10F15 (D621) RPCI-98
            10.F.15 map 42E-43A strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 25 unordered pieces.
ACCESSION   AC007624
VERSION     AC007624.5  GI:6563437
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 167062)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hunmasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Sviraskas,R.R., Wan,K.H., Webster,D.,
            Wooley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167062)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hunmasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Sviraskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Dec 13, 1999 this sequence version replaced gi:5670596.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.berkeley.edu/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 25 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *      1    586: contig of 586 bp in length
          *    587    666: gap of unknown length
          *    667   1636: contig of 970 bp in length
          *   1637   1716: gap of unknown length
          *   1717   3320: contig of 1604 bp in length
          *   3321   3400: gap of unknown length
          *   3401   4834: contig of 1434 bp in length
          *   4835   4914: gap of unknown length
          *   4915   6615: contig of 1701 bp in length
          *   6616   6695: gap of unknown length
          *   6696   8168: contig of 1473 bp in length
          *   8169   8248: gap of unknown length
          *   8249  10093: contig of 1845 bp in length
          *  10094  10173: gap of unknown length
          *  10174  11992: contig of 1819 bp in length
          *  11993  12072: gap of unknown length
          *  12073  14171: contig of 2099 bp in length
          *  14172  14251: gap of unknown length
          *  14252  17553: contig of 3302 bp in length
          *  17554  17633: gap of unknown length
          *  17634  19582: contig of 1949 bp in length
          *  19583  19662: gap of unknown length
          *  19663  22792: contig of 3130 bp in length
          *  22793  22872: gap of unknown length
          *  22873  28892: contig of 6020 bp in length
          *  28893  28972: gap of unknown length
          *  28973  35119: contig of 6147 bp in length
          *  35120  35199: gap of unknown length
          *  35200  40573: contig of 5374 bp in length
```

```
Query Match        10.7%;   Score 150.8;  DB 60;  Length 119597;
Best Local Similarity   63.0%;
Matches 233;  Conservative 0;  Mismatches 137;  Indels 0;  Gaps 0;

          *  83110  83189: gap of unknown length
          *  83190  83720: contig of 4531 bp in length
          *  87721  87800: gap of unknown length
          *  87801  93398: contig of 5598 bp in length
          *  93399  93478: gap of unknown length
          *  93479  97772: contig of 4294 bp in length
          *  97773  97852: gap of unknown length
          *  97853 101858: contig of 4006 bp in length
          * 101859 101938: gap of unknown length
          * 101939 102485: contig of 547 bp in length
          * 102486 102565: gap of unknown length
          * 102566 103236: contig of 671 bp in length
          * 103237 103316: gap of unknown length
          * 103317 103831: contig of 515 bp in length
          * 103832 103911: gap of unknown length
          * 103912 104560: contig of 649 bp in length
          * 104561 104640: gap of unknown length
          * 104641 105200: contig of 561 bp in length
          * 105201 105281: gap of unknown length
          * 105282 105883: contig of 602 bp in length
          * 105884 105963: gap of unknown length
          * 105964 106605: contig of 642 bp in length
          * 106606 106685: gap of unknown length
          * 106686 107172: contig of 487 bp in length
          * 107173 107252: gap of unknown length
          * 107253 107985: contig of 733 bp in length
          * 107986 108065: gap of unknown length
          * 108066 108682: contig of 617 bp in length
          * 108683 108762: gap of unknown length
          * 108763 109459: contig of 697 bp in length
          * 109460 109539: gap of unknown length

Qy   1035 tgactggtatgaagaatatccaactacacctagctcttcttgtttaaatggcttatgta 1094
          ||||||  |||| |||||| ||| ||  | |   ||| || ||||  |||||||||||||
Db  28844 TTACAGGTACGAGGAGTACCCACCACACGCCACCCTCTTATGTCCTAAACGGCTTTATCTA 28785

Qy   1095 ttctttaattgggctgtatgacttaaagaaactgcagggaaaaactcgggaaagaagc 1154
          ||||  |||| ||||| |||||| |||| |  ||| ||||||||| |||  ||||||||
Db  28784 TTCACTGCTGCTGGGTCTTTACGATCTCAACGACACTGCCCCGGGAAAAAATCGCCCGCGAGGC 28725

Qy   1155 gaggtccttgtatgagcgtggcatggaatccttaaagccaatgctccccccttgtacgacac 1214
          ||| | ||| ||||  |||| ||| |  ||   ||||  | || | ||||  |  | ||| |||
Db  28724 TGGAAAGCTTTTCGCGCAGGGCATGCACTCCCCTTAAAAAGATGCTGTTGCTATTTGACAC 28665

Qy   1215 tggctcaggaaccatctatgacctccggcacttcatgcttggccccccaacctggc 1274
          |||||| ||  || |  | |||| | ||||  | || |  ||| ||||| | |||||
Db  28664 TGGCTCCCGGCACCAGCTACGACCTGCGCCATTGTGAGCCGTAGCCTCCTAATCTGGC 28605

Qy   1275 ccgctgggactatcacaccaccaccaccacatcaatcaactgcagctgctagcacattgatga 1334
          ||| ||||||| || ||  ||  |  ||| |  ||| ||  |||  |||  |||||| |||| ||
Db  28604 GCGCTGGGACTACCATGCAACCGACGTGAACCCAGTTACTTCTGTCTAGCCACCATCGACAG 28545

Qy   1335 gtccccaatcctcaaagaatcttgtcaaagaggtggaagaagctaccttaaaggcagccgggc 1394
          |||| |  | ||| ||  |  | || ||||||| ||||| |  ||| |  |||||| |  ||||
Db  28544 CGATCCATTAATTGCACAAAACTGCGGAGGCGCTGGAAGGGCTATATGTTTGGTCGTCGGGC 28485

Qy   1395 aaagcacacac 1404
          |||| |||| |
Db  28484 CAAGCACAAC 28475
```

```
RESULT 11
LOCUS        AE003789     299474 bp    DNA            INV       04-OCT-2000
DEFINITION   Drosophila melanogaster genomic scaffold 142000013386042 section 1
             of 2, complete sequence.
ACCESSION    AE003789  AE002778
VERSION      AE003789.2  GI:10727129
KEYWORDS     HTG.
SOURCE       fruit fly.
  ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
             Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
             Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1  (bases 1 to 299474)
  AUTHORS    Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
             Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
             George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
             Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
             Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
             Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
             Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
             Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
             Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
             Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
             Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
             Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
             Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
             Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
             Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
             Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
             Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
             Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
             Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
             Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
             Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
             Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
             Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
             Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
             McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
             Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
             Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
             Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
             Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
             Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
             Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
             Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
             Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
             Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
             Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
             Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
             Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
             Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE      The genome sequence of Drosophila melanogaster
  JOURNAL    Science 287 (5461), 2185-2195 (2000)
  MEDLINE    20196006
REFERENCE    2  (bases 1 to 299474)
  AUTHORS    Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE      Direct Submission
  JOURNAL    Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
             Rockville, MD, USA
COMMENT      On Oct 9, 2000 this sequence version replaced gi:7302263.
FEATURES             Location/Qualifiers
     source          1..299474
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
     mRNA            join(11901..12167,12272..12335,12376..12335,13327,37670..38167,
                     38262..38332)
                     /gene="1,28"
                     /note="CG9093"
                     /product="CG9093"
                     /db_xref="FLYBASE:FBan0009397"
                     /db_xref="FLYBASE:FBgn0010347"
     gene            <11901..>38332
                     /gene="1,28"
                     /note="CG9397"
                     /map="42A13-42A16"
```

ORIGIN

```
     source          1..167062
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2"
                     /map="42E-43A"
                     /clone="BACR10F15 (D621) RPCI-98 10.F.15"
                     /clone_lib="RPCI-98 (Roswell Park Cancer Institute
                     Drosophila melanogaster BAC library, partial EcoRI in
                     pBACe3.6"

BASE COUNT    45406 a   36877 c   36482 g   46377 t   1920 others
ORIGIN
```

```
Query Match          10.7%;   Score 150.8;  DB 60;   Length 167062;
Best Local Similarity   63.0%;   Pred. No. 2.7e-30;
Matches  233;  Conservative  0;  Mismatches  137;  Indels  0;  Gaps  0;

Qy    1035 tgactggtatggagagaatatccacactagcctcttttgtttaaatggctttatgta 1094
           ||||| ||||| |||||||||||| ||||||||||| |||||||| ||||||||| |||
Db  153555 TTACAGGTACGAGAGAGTACCCACACCACGCACCACCCTCTTPATGTGCCTAAACGGCTTTATCTA 153496

Qy    1095 ttcttaattgggctgtatgacttaaaagaaactgcaggggaaaaactccggaaagaagc 1154
           ||||||||||||||||||||||||||||| ||||| ||||||||||||||||||| |||
Db  153495 TTCACTGCTGTGGGTCTCTACGATCTCAACAGCACTGCCCCGGGAAAAATCGCCGCGAGGC 153436

Qy    1155 gaggtccttgtatgagcgtggcatggaatccctaaagccatgctccccttgtacgacac 1214
           || |||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db  153435 TGGAAAGCTTTTCGCGCAGGCATGCACTCCCTTAAAAAGAATGCTGTTGCTATTTGACAC 153376

Qy    1215 tggctcaggagaacatctatgacctcccggcacttcatggcattgccccaaactggc 1274
           ||||||| || ||||||||||||||||||||| |||||||||||||||||||||||||
Db  153375 TGGCTCCGGCGCCACCAGCTACGACCTGCGGCCCATCTGAGCCGCTAGCTCCTCPAATCTGGC 153316

Qy    1275 ccggctgggactatacacacccccacccaatcaactgcagctcgttagcaccattgatga 1334
           ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db  153315 GCGCTGGGACTACCATGCAAACGACGTGAACCAGTTACTTCTGTTAGCCACCATCGACAG 153256

Qy    1335 gtcccaatcttcaaagaatttgtcaagaggtcgaaagagctaccttaaagcaagccgggc 1394
           ||| |||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  153255 CGATCCATTAATTGCACAAACTGGGAGGGCTCTATATGTTGTTTGGTCCTCGGCGC 153196

Qy    1395 aaagcacaac 1404
           ||||||||||
Db  153195 CAAGGACAAC 153186
```

```
       *        40574     40653: gap of unknown length
       *        40054     50146: contig of 9493 bp in length
       *        50147     50226: gap of unknown length
       *        50227     56630: contig of 6404 bp in length
       *        56631     56710: gap of unknown length
       *        56711     65243: contig of 8533 bp in length
       *        65244     65323: gap of unknown length
       *        65324     78848: contig of 13525 bp in length
       *        78849     78928: gap of unknown length
       *        78929    106751: contig of 27823 bp in length
       *       106752    106831: gap of unknown length
       *       106832    130045: contig of 23214 bp in length
       *       130046    130125: gap of unknown length
       *       130126    165293: contig of 35168 bp in length
       *       165294    165373: gap of unknown length
       *       165374    165599: contig of 226 bp in length
       *       165600    165679: gap of unknown length
       *       165680    166287: contig of 608 bp in length
       *       166288    166367: gap of unknown length
       *       166367    167062: contig of 695 bp in length.
```

```
CDS             /db_xref="FLYBASE:FBan0009397"
                /db_xref="FLYBASE:FBgn0010347"
                12648. .13301
                /gene="1.28"
                /note="1.28 gene product; Nucleotide sequence of the
                Celera sequence differs from the published sequence for
                this transcript"
                /codon_start=1
                /db_xref="FLYBASE:FBan0009397"
                /db_xref="FLYBASE:FBgn0010347"
                /protein_id="AAF57355.1"
                /db_xref="GI:7302264"

mRNA            complement(join(<109514. .109577,109653. .109695,
                10991. .110233,110341. .110536. .>110587))
                /gene="CG15233"
                /product="CT35170"
                /db_xref="FLYBASE:FBan0015233"
                /db_xref="FLYBASE:FBgn0033076"
                complement(<109514. .>110587)

gene            /gene="CG15233"
                /map="42B1-42B2"
                /db_xref="FLYBASE:FBan0015233"
                /db_xref="FLYBASE:FBgn0033076"
                complement(join(109514. .109577,109653. .109695,
                10991. .110233,110341. .110397,110536. .110587))
                /gene="CG15233 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0015233"
                /db_xref="FLYBASE:FBgn0033076"
                /protein_id="AAF57356.1"
                /db_xref="GI:7302267"
                /translation="MSSGALNTLAIVLLLMPGAYVHQWVVPEEDYHSDLKGYIETFCC
                LYSPSIGCOLALNPGIFNEKKLPVRRCKDCIEHCQCSNKSTPPLKPRMVVSLLGMFAM
                LIESIYITQKVWAGCALGPGKGLEPRYGRRHGHNAIVIVAMEKSSPPPQE"

CDS             complement(<116891. .>117118)
                /gene="CG15234"
                /product="CT35171"
                /db_xref="FLYBASE:FBan0015234"
                /db_xref="FLYBASE:FBgn0040673"
                /evidence=not_experimental
                complement(<116891. .>117118)
                /gene="CG15234"
                /map="42B1-42B2"
                /db_xref="FLYBASE:FBan0015234"
                /db_xref="FLYBASE:FBgn0040673"
                /evidence=not_experimental
                complement(116891. .117118)
                /gene="CG15234"
                /note="CG15234 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0015234"
                /db_xref="FLYBASE:FBgn0040673"
                /protein_id="CT40117"

gene            join(123398. .123479,123714. .125350,125546. .127154,
                127220. .127781,127850. .127997,128173. .128350,
                128412. .128862)
                /gene="CG9403"
                /product="CT9101"
                /evidence=not_experimental
                /db_xref="FLYBASE:FBan0009403"
                /db_xref="FLYBASE:FBgn0033077"
                /translation="MFAHALWLRLAPSSTKTRGPGPGRGPGPGPCAGLGLKTGTGTGRGPEV
                WPRVKEGQGQGLSPAAIDRVEQIEKQQKKSA"
                join(123398. .123479,123714. .125350,125546. .127154,
                127220. .127781,127850. .127997,128173. .128350,
                128412. .128862)
                /gene="CG9403"
                /product="CG9403"
                /db_xref="FLYBASE:FBan0009403"
                /db_xref="FLYBASE:FBgn0033077"
                <123398. .>128862
                /gene="CG9403"
                /map="42B2-42B2"
                /db_xref="FLYBASE:FBan0009403"
```

```
CDS             /db_xref="FLYBASE:FBgn0033077"
                join(123838. .125350,125546. .127154,127220. .127781,
                127850. .127997,128173. .128350,128412. .128862)
                /gene="CG9403"
                /note="CG9403 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0009403"
                /db_xref="FLYBASE:FBgn0033077"
                /protein_id="AAF57358.1"
                /db_xref="GI:7302267"
                /translation="MPPHILGSASGSSSTAAAASPPISSASQTPPSPAIHHSLSQHNA
                TSAMSKPTTPQRHGNPSLLPTLQWPWNTSLGSTSTAVPAASANKNKRTAACSGAALGL
                SGEGSCLIPGGGAASAKARSDLPGSFPDASKRLKVAAMEESQTKITCFFKSQMKPSPGG
                GKLSPQPCQOSNPANTLIMSTPATTASLNKYFNILSQLKEQAQQQOOASKTLPT
                PAPVAPVPESAPPVAPVVVTPSPSLPVPALKKIERSNKQPAKIAQVAPNLRKTPSSCSS
                GSSSSSSSSSNGSSNTGKSPTKHVHAIAPRTPEMKQOQOOOKKAAMYYRPPVTSTALKQ
                KOISPPAPVAAHTPAPALNPTPAPANPTLYQLPVQLPNLVQLPQLAAANIMQLN
                NVAKAAVAAAATNNAAAQOAAQAAQAYFLNGTVRKLQOVTTATTTATTATTATAAA
                PAGNPFGLLNLATAGNFEPGLPNANGPEIEAIPGACSYLIHHOLLLARQNMMSLNETMA
                SNSCGNNNFVNPLNNQOVGLKDNKLYIVNSAEYLGIYLMSLQIALNNOHOQQOIISAT
                QPPLAATNNNSNTATQPPPLAATNTNNSNTAFNNSTASNINHRASNNLSNINNTAI
                QPQRAIVKPPLHSSTOPPPLVTVISSMEPACSPAAASPAAKRSLPAAKPYQKRLTKGR
                VGTAPIIATPTPPTPPLVPTSATKELGQLRKSTGCTTGTPTGTPPTGTFTPTPLVSIAPSKLTP
                TLSVSKQGPTMKLANSAPDLFDLVKNSKLVARVSQPLTPLPFSSPSSPSSNGSHGSHGM
                RSSPALTSTNSCTLSAFSKIKYETTELASQTGSITLSSIIPPISLIKPQSFAGLQPKREP
                ESETDLIKHDLLPDCTNSNSNSSNSCSSSTYSHSVSSAADLSLEASTPAPSPSPSASPS
                GLGSPSPAASNLSASSRRAASQTDMLSELVTSSCISSGGDCSQATDSPMPALPLAK
                SEDATTPISTVSGGSSSSGSSNYDEEDDKSVASLETHOTHRKLRLDLPTPEGGIGGSLSN
                SESSNSIADAISSKSASVGPPTAASSASASDSNSLASNAPSPAASPEDCSAAPSPA
                CSASTTGSIPPSTVVDIAMVEATKSLPFKSAISPILSOPKIIRFPAPAGASGKGKRH
                DGVCYWDKCNKHHESNSKLLDHMOTHHVNTQTGPFACLWVGCKVVNKESCSRRWIERH
                VLSHGGSKQFKCIVEGCGLRFCSQLALQKHVNNHFNATDNARESTSKRTSDPPVPKQL
                RKNGKKLRYRRQPFSARMFDFFDTGIMEGLQHRLRQISTLLTNGAQAIEFQOQCMRRR
                NSQGGYECFVRWSPREIISDEWLPEECLNRTRQHTKVLHILKHKOMRPAEKTRVDSLLSTAF
                RLRYDSHLFADDYNVNEQQVGEASGSDCDEDGDGDEGDDEEDLEQDEEEGSSSSRS
                ASCETVSSYQQVLSIARLQMQQORRKHPRKPRKVTPPAREKLVPTDALVPI"
                complement(join(136536. .137283,137341. .137530,
                141637. .141979))

mRNA            /product="CT12409"
                /gene="Vha16"
                /db_xref="FLYBASE:FBan0031161"
                /db_xref="FLYBASE:FBgn0004145"
                complement(join(136536. .137283,137341. .137530,
                141637. .141750,141934. .142026))
                /gene="Vha16"
                /product="CT40117"
                /db_xref="FLYBASE:FBan0031161"
                /db_xref="FLYBASE:FBgn0004145"
                complement(join(136536. .137283,137341. .137530,
                141637. .141750,142605. .142690)))
                /gene="Vha16"
```

```
Query Match        10.7%;  Score 150.8;  DB 4;  Length 299474;
Best Local Similarity   63.0%;  Pred. No. 2.7e-30;
Matches  233;  Conservative  0;  Mismatches  137;  Indels  0;  Gaps  0;

Qy    1035 tgactggtatgaagaatatccaactacacctagctcttttgtttaaatggcttatgta 1094
           ||| ||||  |||| |||| || ||||||| |||||  |||| ||||||||| |||||
Db  180285 TTACAGGTACGAGGAGTACCCCACCACGCCACCCCTCTTATGTGCTAAACGGCTTTATCTA 180226

Qy    1095 ttcttaattgggctgtatgacttaaaagaaactgcaggggaaaaagaagc 1154
           || |  |||| || ||| || ||| || |||| || |||||| |||||||
Db  180225 TTCACTCGCTGGGTCTCTACGATCTCAACAGCACACTGCCCCGGGAAAAATCGCCCGCGAGGC 180166

Qy    1155 gaggtccttgtatgagcgtggcatggaatccttaaaagccatgctcccccttgtacgacac 1214
           ||| | ||||  ||  |||| || || |||||||| ||||||||| | || |  | ||||
Db  180165 TGGAAAGCTTTTCGCCGCAGGGCATGCAACTCCCTTAAAAAGATGCGTTGCTTATTGACAC 180106

Qy    1215 tggctcaggaaccacatctatgacctcgggcactcatgcttggccaccaacctggc 1274
           ||||   |  |||||| ||  |||||||| |  |||||| || || |||| ||||||
Db  180105 TGGCTCCGGCACCAGCTACGACCTGCGCCATCTGACCCTGGCGTAGCTCCTAATCTGGC 180046

Qy    1275 ccgctgggactatcaccacccacaatcaactgcagctgcttagcacaccattgatga 1334
```

```
Db  180045 GCGCTGGGACTACCATGCCAACGACACGTGACCACCAGTTACTTCTGTTGTGTTAGCCACATCGCACAG 179986
              |||||| |||  |||  || ||   ||| || ||||||  ||||||| ||| ||| ||||||||| ||
Qy    1335 gtccccaatcttcaaagaaatttgtcaaagaggtggaagagagctacctcaaggcagccggggc 1394

Db  179985 CGATCCATTAATTGCACAAACTGCGGAGCCGCTGGAAGGGCTATATGTTGGTCGTCGGGC 179926

Qy    1395 aaagcacaac 1404
             ||||||||||
Db  179925 CAAGCACAAC 179916
```

RESULT  12
CEB0285

LOCUS       CEB0285       41397 bp    DNA       INV       23-OCT-2000
DEFINITION  Caenorhabditis elegans cosmid B0285, complete sequence.
ACCESSION   Z34533
VERSION     Z34533.1  GI:506879
KEYWORDS    HTG; Aminopeptidase; CDC2-like protein kinase; Choline kinase;
            Glycine-rich domain; Mucin; transfer RNA; tRNA-Leu.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 41397)
  AUTHORS   none.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  REMARK    Erratum:[published errata appear in Science 1999 Jan
            1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
            3;285(5433):1493]]
REFERENCE   2 (bases 1 to 41397)
  AUTHORS   Sulston,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1994) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            currently available information.
            Coding sequences below are predicted from Genefinder (P. Green, U.
            Washington), and other
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is not the entire insert of clone B0285.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone B0285 is at 25507 in
            sequence AL024499.
            The start of this sequence (1. .104) overlaps with the end of
            sequence AL024499.
            The end of this sequence (41294. .41397) overlaps with the start of
            sequence Z30973.
            Predicted proteins B0285.8, B0285.9 and B0285.10 are 60-708
            identical at the amino acid level.
            For a graphical representation of this sequence and its analysis
            see:- http://webace.sanger.ac.uk/perl/ace/elegans/seq/sequence?
            name=B0285
            [000905 d1] Merged B0285.1 and B0285.2 based on EST data.
FEATURES             Location/Qualifiers
     source          1. .41397
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="B0285"
     gene            join(2715. .2908,3445. .3642,5546. .5672,6107. .6220,
                     6278. .6498,6852. .6997,7050. .7339,7398. .7471,7733. .7888,
```

```
                     8139. .8333,8382. .8532,8827. .9015)
                     /gene="B0285.1"
     CDS             join(2715. .2908,3445. .3642,5546. .5672,6107. .6220,
                     6278. .6498,6852. .6997,7050. .7339,7398. .7471,7733. .7888,
                     8139. .8333,8382. .8532,8827. .9015)
                     /gene="B0285.1"
                     /codon_start=1
                     /protein_id="CAA84302.2"
                     /db_xref="GI:1024555"
                     /note="contains similarity to Pfam domain PF00069
                     (Eukaryotic protein kinase domain), Score=302.4,
                     E-value=1.7e-87, N=1"
                     /translation="MEISPGSSTHRDRKGSYGHRERTRSHSGSPSRFYSKDRGSSR
                     QGVRPRDRDSISPQYKQRNWSRGGGGGGGCGRDRGRNDFSYRKGKDINKRDKRSRS
                     RSRHRSPKRSGSSKSKRRNSGSSGSSDLMDTSLMSELKHGDYGSSSKSKRSRKRR
                     KHSSSSSSGGEAMDLPVSSNGMNVTAIPPPSNINFPQFMFSQPPPLPNSQFM
                     TPPRPPPAPFSIPPPSVDIHFAATASFSLSSIPPPPQTDGGASSSKRQDLPMPPD
                     SKRIATRPVITTRRGHATNRPSDSDGTVLHTHTMLDQIGESTGQVKAVNNLTG
                     EQVALKRVRLENEKGFPTIAIREIKILRQLHHKNIVRLMDIVIDISMDELKRTRAN
                     FYLVFEYVDHDLIGLLESKELVDFNKDQICLSFKQLLEGLAYIHNTGFLHRDIKCSNI
                     LVNNGELKIADLGLARLWEKESRLYTNKVWRPPELLIGDERYGFPIADVWSVGCI
                     MLGELFTRKPLFNGNNEFGQLELISKVCGSPNVDWPELTELVGNNTFRMKRTYQRRI
                     REEFEHIMPREAVDLLDKMLTLNPEKRISAKEALNHPWIRSLEHTTVQPLKPQHQDC
                     HEMWSKKQKRSARLGRQAESGSGSSHSIRATSHPRGKAGDELEIKISDFIFQFSENK
                     VN"
     gene            complement(join(11398. .11451,11504. .11608,12102. .12397,
                     12467. .12512))
                     /gene="B0285.3"
     CDS             complement(join(11398. .11451,11504. .11608,12102. .12397,
                     12467. .12512))
                     /gene="B0285.3"
                     /note="contains glycine-rich domain
                     /codon_start=1
                     /protein_id="CAA84296.1"
                     /db_xref="GI:3873727"
                     /db_xref="SWISS-PROT:P46553"
                     /translation="MNYGNNGGQOWOPRGNNNWRPGGRGRGGGGWKGNGNDRGGYRGGS
                     SNDGGWSNFRDDGNLNEFASPYQANRGNFRGNYRGGGGGGGNQONNYGNRGRGGFDR
                     GRGGGGGNKNFGPIDANKCIIPSMWNNPWDALEREKEYEQEYGVAATEKVTDPSPSPSS
                     AEETTA"
     gene            join(12779. .12919,12968. .13145,13248. .13412,13465. .13603,
                     13901. .14107,14164. .14250,14306. .14390)
                     /gene="B0285.4"
     CDS             join(12779. .12919,12968. .13145,13248. .13412,13465. .13603,
                     13901. .14107,14164. .14250,14306. .14390)
                     /gene="B0285.4"
                     /note="cDNA EST yk173g10.3 comes from this gene
                     cDNA EST yk173g10.5 comes from this gene
                     cDNA EST yk217f7.3 comes from this gene
                     cDNA EST yk217f7.5 comes from this gene"
                     /codon_start=1
                     /protein_id="CAA84295.1"
                     /db_xref="GI:3873726"
                     /db_xref="SWISS-PROT:P46554"
                     /translation="MDSEAVSSDSHVAAAIAIRRRSNSVSDDYSVQRTNDDATQCKYF
                     ATQKGYWKDEFISRFANSSSNVSEARRFPEISMGYKWARTAAIEKYVRDFLNEFDGNAQ
                     VVSLGCGPDTLFWRLVSSGAKLVKVEVDFSSVTSKKIRHIHKPIGPNSVDLKKSFES
                     DAVVSHHADLHAGNYHLIGADLRQANELDQKLATCQLSHDIPTIFIAECVLVIMSADS
                     STALLKGTVSQFKQPAFVNYEQFRTSDAFTKYMEQNLGDRGIQLHGLEMCESAEKQEE
                     RFRNAGFREVKVWMDMNQIFNNFLDQKEVSRIFEIEMLDEMLLQQLFAHVCVVSARI"
     gene            join(14571. .14615,14657. .14818,15191. .15485,15715. .15810,
                     15860. .16018,16144. .16259,16388. .16658,17070. .17288,
                     17776. .18311)
                     /gene="B0285.5"
     CDS             join(14571. .14615,14657. .14818,15191. .15485,15715. .15810,
                     15860. .16018,16144. .16259,16388. .16658,17070. .17288,
                     17776. .18311)
                     /gene="B0285.5"
                     /codon_start=1
                     /note="cDNA EST yk96a2.5 comes from this gene"
                     /protein_id="CAA84297.1"
```

```
/db_xref="GI:3873728"
/db_xref="SWISS-PROT:P45555"
/translation="MVLVSLKPFNIFSLKPMKCLRWRSNRHRIYLLVACGALFLLNRH
LTQEESRIDEDEELIQVDVNEDDKRVEPFGSIESKCIIADNRSKMCKWKDEEDVYF
PVSILKRFDMTGKLGKDGSSFPELYTSYAKMRSPDSTYDPLGPFGHFSTSVETRDRV
RCVSAKTDVPMSTOWDPIPYYPIQISOYGLQHYSRMKLDSISNKSEASPKDDVILGV
NSKEKGAAGMHETTERLFNDEQMGWSAVVNISAGANLANAGAYYILDKSPDLHYISFD
WKPYEANSSFVLAKMKQDLLVLINVYSEGNGKCVWQEEERISDDYIVQKPKKDGQ
VSYSYSYIGNSPIGEWSTVRDLLVDVARLSSGDNRKKDDNVLHAGDLRLVSLGFR
GELTVKQKITQRREOHSHAFYAAWRWLVKNQNDRGGWSVPVERSIAERKLVLPPGWHS
AMAQGHIGISVLTRAFKHFNDEKYILKSAAKALKLFKINSSDGVRGEFFGNIWYEEYPT
TPGSEVLNGFYLYSLGLYDLSOLELMIDENDETMRAKIOEAQELYSAGVRSLKQLLPL
YDTGSGTTIYDLRHVALGTAPNLARNDYHAVHVYLLKWIAGIEKDEVLSKFADRWIGYA
YGKRAKHN"
/protein_id="CAA84299.1"
/codon_start=1
/note="threonine-rich domain at N terminus; short region
of similarity with aminopeptidases, contains similarity to
Pfam domain: PF01433 (Peptidase family M1), Score=130.2,
```

```
/db_xref="GI:3873730"
/db_xref="SWISS-PROT:P46556"
/translation="MFFESTEGKCAPVILTMSCYWVAEVVPLAVTSFIPMIALPPLGI
VSIKEVAPKYFADTINIVFFNSIMLSLAVEECOLHRRIALKMLTVGTRPHWLMAGFMI
ITSFTSLWISDTACCALMAPIAYALLEEIMIPKMRPEEKENEIEVMKIFDKEDPEKE
KKKLDFSRLSVRDRGICKCMMLLVAHASLIGGTGTINSTGPNLIFRDNIEKNFPNEDH
GISYLSWMAFAIPPMIFYMEFSSWFIVQLQFLGPRHLMGMFREPTETKOEEEVAKRAV
WKSYDOLGPWTWAEKSTLVIFVLAVLSWVSSDPKVIPGWSDLFRGYVTDSCSGLVAV
FLFFIMPKKKPDFRIFRKDKSRPSVRQEPLIDWDCVRRREPWSIILLLGAGFAISDAV
RVSGLSSLIACSLNSTHSIKMPFFVWQIILSIVVVWMIESTNSATASIFIPISRKMAE
AVGAHPLYFSTPTAIGPSFSFMLPMATPANAIVTETKTIRMIDMVSCGVPLNIFCIAI
TAINMNTWAFWLFNMGTYDYDALRIHATNMTGNSSOCF"
/gene="B025.7"
/gene="B025.7"
```

```
Qy    1032  acatgactggtgtgaagaatatccaactacacctagctcttcttttgttttaaatgctttat  1091
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17909  AAATATTTGGTATGAGGAAATATCCAACAACACCGGGATCTTTGTGTCCAATGGATTCCT  17968

Qy    1092  gtattcttaattgggctgtatgactaaaagaaaactgcagggggaaaactcggga-----  1147
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17969  TTATTCACTTATTGGACTTTATGATCTTTCACAATTGGAATTGATGATTGCATGGAAAATGA  18028

Qy    1148  ------------------aagaagcgaggtccttgtatggacgtggcatggaatc  1184
                              |||||||||||||||||||||||||||||||||
Db   18029  TGAAACAATGCGGGCGCAAAATTCAAGAGGCACAAGAGAATTATATTCAGACAGTGATC  18088

Qy    1185  ccttaaagcaatgctccccttgtacgacacatggctccaggagaccatcctatgaccccgggca  1244
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18089  ACTGAAACAACTCTCCCCTTTATATGATGACTGGCTAGTGGAACGAATTTATGAGTCTTTCGTCA  18148

Qy    1245  cttcatgcttggcattgccccaactgtggcccgctgggactatcacacacccaatcaa  1304
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18149  CGTTGCTTTGGGAACTGCACCGGAATCTTGCAAGATGGGAATTATCATGCTGTACATGTGTA  18208

Qy    1305  tcaactgcagctgcttagcaccattgatgagtccccaaatcttcaaagaagatttgtcaagag  1364
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18209  CTTATTGAGAATGGATTGCCTGGAATCAAAAGATGAAGTTTTGACTGATAAAAAACAGCAGATCG  18268

Qy    1365  gtgggaagagctaccttaaaggcagccgggcaaaagcacaa  1403
            |||||||||||||||||||||||||||||||||||||
Db   18269  ATGGATTGGGATACGGCGTTATGGGAAAACGGGCAAAAACATAA  18307
```

```
gene  join(21833..21850,22021..21170,22213..22328,22611..22646,
          22695..22853,22902..22945,22990..23098,23150..23377,
          23585..23781,23825..23946,23993..24295,25051..25206)
      /gene="B025.6"

CDS   join(21833..21850,22021..21170,22213..22328,22611..22646,
          22695..22853,22902..22945,22990..23098,23150..23377,
          23585..23781,23825..23946,23993..24295,25051..25206)
      /note="similar to C. elegans R107.1, K08B5.2; localised
      similarity to yeast URF YCR37C and rat mucin (PIR:S26784),
      contains similarity to Pfam domain: PF00939
      (Sodium:sulfate symporter transmembrane region),
      Score=153.0, E-value=2.4e-43, N=2"
      /codon_start=1

gene  join(26002..26439,27474..27738,30045..30685,31287..31397,
          32550..32792,32888..33229,33357..33515,33765..33911)
      /gene="B025.7"

CDS   join(26002..26439,27474..27738,30045..30685,31287..31397,
          32550..32792,32888..33229,33357..33515,33765..33911)
      /gene="B025.7"
```

```
RESULT  13
LOCUS       AL160170     156257 bp     DNA     HTG       20-JAN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-234F20, *** SEQUENCING IN
            PROGRESS ***, 16 unordered pieces.
ACCESSION   AL160170
VERSION     AL160170.4  GI:9863586
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156257)
AUTHORS     Plumb,B.
TITLE       Direct Submission
JOURNAL     Submitted (19-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 21, 2000 this sequence version replaced gi:9187197.

            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------------- Project Information ------------------
            Center project name: bA234F20
            -------------------- Summary Statistics --------------------
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 147428 bases at least Q40
            Consensus quality: 150614 bases at least Q30
            Consensus quality: 152462 bases at least Q20
            Insert size: 154757; sum-of-contigs
            Insert size: 151975; 15.3% error; agarose-fp
            Quality coverage: 3.45x in Q20 bases; sum-of-contigs Quality
            coverage: 3.61x in Q20 bases; agarose-fp

            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
```

```
cDNA  EST  cm04a2   comes from this gene
cDNA  EST  EMBL:TO2449  comes from this gene
cDNA  EST  EMBL:TO0914  comes from this gene
cDNA  EST  EMBL:TO2161  comes from this gene
cDNA  EST  yk51d6.5   comes from this gene
cDNA  EST  yk51d6.3   comes from this gene
cDNA  EST  EMBL:M75802  comes from this gene
cDNA  EST  yk485a8.3  comes from this gene
cDNA  EST  yk485a8.5  comes from this gene
cDNA  EST  yk327e7.3  comes from this gene
cDNA  EST  yk327e7.5  comes from this gene
cDNA  EST  yk345a2.3  comes from this gene
cDNA  EST  yk345a2.5  comes from this gene
cDNA  EST  yk266f2.3  comes from this gene
cDNA  EST  yk24a66.3  comes from this gene
cDNA  EST  yk24a66.5  comes from this gene
cDNA  EST  yk23016.3  comes from this gene
```

```
Query Match     5.4%;   Score 75.8;   DB 5;   Length 41397;
Best Local Similarity  52.6%;   Pred. No. 9.9e-10;
Matches  210;  Conservative  0;  Mismatches  162;  Indels  27;  Gaps  1;
E-value=1.6e-05, N=1
```

```
* as soon as it is available and the accession number will
* be preserved.

FEATURES             Location/Qualifiers
source               1..156257
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-234F20"
                     /clone_lib="RPCI-11.1"
misc_feature         1..2820
                     /note="assembly_fragment:00664
                     fragment_chain:1"
misc_feature         2821..2920: gap of 100 bp in length
                     2921..4935
                     /note="assembly_fragment:00425
                     fragment_chain:1"
misc_feature         4936..5035: gap of 100 bp in length
                     5036..8844
                     /note="assembly_fragment:00186
                     fragment_chain:1"
misc_feature         8845..8944: gap of 100 bp in length
                     8945..26698
                     /note="assembly_fragment:01136
                     fragment_chain:1"
misc_feature         26699..26798: gap of 100 bp in length
                     26799..50840
                     /note="assembly_fragment:00193
                     fragment_chain:2"
misc_feature         50841..50940: gap of 100 bp in length
                     50941..53114
                     /note="assembly_fragment:00266
                     fragment_chain:2"
misc_feature         53115..53214: gap of 100 bp in length
                     53215..64925
                     /note="assembly_fragment:00274
                     fragment_chain:2"
misc_feature         64926..65025: gap of 100 bp in length
                     65026..102551
                     /note="assembly_fragment:00554
                     fragment_chain:3"
misc_feature         102552..102651: gap of 100 bp in length
                     102652..113047
                     /note="assembly_fragment:00508
                     fragment_chain:3"
misc_feature         113048..113147: gap of 100 bp in length
                     113148..126763
                     /note="assembly_fragment:00425
                     fragment_chain:4"
misc_feature         126764..126863: gap of 100 bp in length
                     126864..129456
                     /note="assembly_fragment:00596
                     fragment_chain:4"
misc_feature         129457..129556: gap of 100 bp in length
                     129557..131856
                     /note="assembly_fragment:01019
                     fragment_chain:4"
misc_feature         131857..131956: gap of 100 bp in length
                     131957..142080
                     /note="assembly_fragment:01103
                     fragment_chain:4"
misc_feature         142081..142180: gap of 100 bp in length
                     142181..146194
                     /note="assembly_fragment:00759
                     fragment_chain:4"
misc_feature         146195..146294: gap of 100 bp in length
                     146295..148386
                     /note="assembly_fragment:00059
                     fragment_chain:4"
misc_feature         148387..148486: gap of 100 bp in length
                     148487..156257
                     /note="assembly_fragment:00456
                     fragment_chain:4"
BASE COUNT           48192 a    28966 c    29334 g    48244 t    1521 others
ORIGIN
                     vector_side:right"
                     clone_end:T7
```

```
Qy   898   caagccattctacattagtcaggcgcctatctcttaacaaaagaccatatattcctcaat   957
           ||| | |||||| ||||| |  || |||| |||||||| ||||||| ||| || ||||||
Db   63599 CAAATTATTAAACAAAAAGTCATACTTTGAAACAAAACTTAATGATATAAAATCCTCACC   63540

Qy   958   tcagctttaaggggcaacagcccccttacaaagttctgtcagagcagcatggagtcaaggct   1017
           ||   ||| ||||| | ||||||||| ||||||| | |||| || || ||||||  ||||
Db   63539 AAATCTACAACAGATCATTTTGCAGTTTACTCAACTTTGTTGTAAAAAAATGTAGAAAAAAT   63480

Qy   1018  gtgtttatgaataaacatgactggtatgaagaatatccaactacaccctagctctttgtt   1077
           | ||| | || ||| ||| |||| ||| | ||| ||  ||||||| | |||||||| |||
Db   63479 GTTTTATAGTGTTAACAGCTCTTATGTACAGACAACAGAATCACTAATAATAATTAATTTCTT   63420

Qy   1078  ttaaatggcttttatgtatcttttaattgggctgtatgacttaaaagaaactgcagggaa   1137
           ||| |||| | | | ||||||  |  ||||||| | ||  ||||| | ||||| || ||
Db   63419 TCACATATTTTATATTTCTTAATAATTTTCAAACTTCTACAATGATTATATAATATTG   63360

Qy   1138  aaactcgggaaagaa   1152
           ||||| |||| ||||
Db   63359 CAAATTTAAAAGAA   63345
```

```
RESULT  14
AL158841/c
LOCUS        AL158841          167721 bp    DNA                    PRI        08-NOV-2000
DEFINITION   Human DNA sequence from clone RP11-183M13 on chromosome 1, complete
             sequence.
ACCESSION    AL158841
VERSION      AL158841.6  GI:11136902
KEYWORDS     HTG.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 167721)
AUTHORS      Hall,R.
TITLE        Direct Submission
JOURNAL      Submitted (07-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
COMMENT      On Nov 10, 2000 this sequence version replaced gi:11121381.
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the sequence submission
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence has been finished according to sequence map criteria
             as follows. An attempt is made to resolve all sequencing problems,
             such as compressions and repeats, but not necessarily within known
             annotated human repeat sequence elements (e.g. Alu). Where the
             sequence is ambiguous, there is an annotation using the 'unsure'
             feature key.
             The following abbreviations are used to associate primary accession
```

numbers given in the feature table with their source source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-183M13 is from the library RPCI-11.1 constructed at the
Roswell Park Cancer Institute by the group of Pieter de Jong. For
further details see http://bacpac.med.buffalo.edu/
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-183M13 The true
left end of clone RP5-831021 is at 111240 in this sequence.

FEATURES             Location/Qualifiers
source               1. 167721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone_lib="RPCI-11.1"
                     /clone="RP11-183M13"
repeat_region        166. .337
                     /note="L2 repeat: matches 2017. .2195 of consensus"
repeat_region        350. .568
                     /note="MLT1B repeat: matches 1. .223 of consensus"
repeat_region        570. .1016
                     /note="L1M2 repeat: matches 5168. .5617 of consensus"
repeat_region        1119. .1184
                     /note="L2 repeat: matches 2210. .2276 of consensus"
repeat_region        1189. .1327
                     /note="MLT1B repeat: matches 212. .389 of consensus"
repeat_region        1328. .1393
                     /note="33 copies 2 mer tt 72% conserved"
repeat_region        1394. .2645
                     /note="L1PA2 repeat: matches 4895. .6146 of consensus"
repeat_region        2679. .2702
                     /note="L1P repeat: matches 4898. .4921 of consensus"
repeat_region        2727. .3230
                     /note="L1MA9 repeat: matches 5724. .6270 of consensus"
repeat_region        3290. .3329
                     /note="L2 repeat: matches 2575. .2614 of consensus"
repeat_region        4634. .4810
                     /note="MIR repeat: matches 25. .212 of consensus"
repeat_region        5377. .5562
                     /note="MIR repeat: matches 60. .262 of consensus"
repeat_region        7088. .7654
                     /note="L2 repeat: matches 2088. .2750 of consensus"
repeat_region        7653. .7704
                     /note="L2 repeat: matches 2698. .2750 of consensus"
repeat_region        7895. .8052
                     /note="MER58B repeat: matches 187. .341 of consensus"
repeat_region        9172. .9925
                     /note="Tigger2a repeat: matches 1. .289 of consensus"
repeat_region        12563. .12913
                     /note="THE1A repeat: matches 1. .354 of consensus"
repeat_region        12914. .14507
                     /note="L1MII repeat: matches 41. .410 of consensus"
repeat_region        14508. .14850
                     /note="THE1A-internal repeat: matches 1. .1580 of consensus"
repeat_region        14851. .15021
                     /note="THE1A repeat: matches 1. .354 of consensus"
repeat_region        16263. .16372
                     /note="Tigger2a repeat: matches 289. .434 of consensus"
repeat_region        16720. .17573
                     /note="L2 repeat: matches 2627. .2746 of consensus"
repeat_region        18729. .18975
                     /note="THE1C repeat: matches 1. .371 of consensus"
repeat_region        19521. .19829
                     /note="TIGGER1 repeat: matches 2172. .2418 of consensus"

repeat_region        20023. .20996
                     /note="AluY repeat: matches 6. .311 of consensus"
repeat_region        21916. .21981
                     /note="L1MEC repeat: matches 274. .1293 of consensus"
repeat_region        22157. .22563
                     /note="L2 repeat: matches 2683. .2749 of consensus"
repeat_region        24729. .25549
                     /note="MSTD repeat: matches 1. .394 of consensus"
repeat_region        26030. .26119
                     /note="L1MC5 repeat: matches 6676. .7531 of consensus"
repeat_region        26692. .26981
                     /note="L2 repeat: matches 2627. .2710 of consensus"
repeat_region        29249. .35342
                     /note="AluSq repeat: matches 1. .294 of consensus"
repeat_region        37018. .37075
                     /note="L1PA6 repeat: matches 4. .6143 of consensus"
repeat_region        37670. .37876
                     /note="29 copies 2 mer ta 79% conserved"
repeat_region        39156. .39294
                     /note="MIR repeat: matches 56. .250 of consensus"
repeat_region        39833. .39898
                     /note="L1M4 repeat: matches 3246. .3378 of consensus"
repeat_region        40670. .40894
                     /note="MLT1G repeat: matches 459. .566 of consensus"
repeat_region        40948. .41068
                     /note="HERVL repeat: matches 2313. .2543 of consensus"
repeat_region        43793. .43955
                     /note="FLAM_A repeat: matches 1. .121 of consensus"
repeat_region        43992. .45055
                     /note="MERVL repeat: matches 5073. .5239 of consensus"
repeat_region        45062. .45407
                     /note="HERVL repeat: matches 3920. .5055 of consensus"
repeat_region        45433. .45623
                     /note="MERVL repeat: matches 2266. .2621 of consensus"
repeat_region        45834. .46449
                     /note="L1M4 repeat: matches 2636. .3033 of consensus"
repeat_region        46611. .46714
                     /note="L1M2 repeat: matches 5063. .5678 of consensus"
repeat_region        46771. .47268
                     /note="52 copies 2 mer at 63% conserved"
repeat_region        47298. .47513
                     /note="L1MA9 repeat: matches 5774. .6288 of consensus"
repeat_region        47535. .47797
                     /note="L1M4 repeat: matches 3035. .3259 of consensus"
repeat_region        47783. .48349
                     /note="MER4D repeat: matches 1. .251 of consensus"
repeat_region        48356. .49070
                     /note="MER4D repeat: matches 396. .950 of consensus"
repeat_region        49251. .49608
                     /note="MER50 repeat: matches 1. .711 of consensus"
repeat_region        49667. .49757
                     /note="L1M4 repeat: matches 5189. .5550 of consensus"
repeat_region        51051. .51304
                     /note="L1M4 repeat: matches 5557. .5647 of consensus"
repeat_region        52088. .52150
                     /note="127 copies 2 mer tt 56% conserved"
repeat_region        52980. .53147
                     /note="L2 repeat: matches 2687. .2750 of consensus"
repeat_region        53870. .53960
                     /note="MIR repeat: matches 22. .188 of consensus"
repeat_region        54155. .54580
                     /note="MIR repeat: matches 50. .140 of consensus"
repeat_region        54647. .54826
                     /note="L2 repeat: matches 445. .881 of consensus"
repeat_region        55219. .56488
                     /note="MIR repeat: matches 8. .197 of consensus"
repeat_region        57097. .57418
                     /note="L1PA7 repeat: matches 4863. .6139 of consensus"
repeat_region        57424. .57638
                     /note="L1MA8 repeat: matches 5975. .6291 of consensus"
repeat_region        57634. .58012
                     /note="L1MC4 repeat: matches 7168. .7388 of consensus"
                     /note="L1MC4 repeat: matches 6605. .7008 of consensus"

```
repeat_region    58014..58121
                 /note="54 copies 2 mer ta 62% conserved"
repeat_region    59702..59794
                 /note="L2 repeat: matches 2618..2710 of consensus"
repeat_region    59926..60232
                 /note="AluSg repeat: matches 1..308 of consensus"
repeat_region    6074..6874
                 /note="FLAM_A repeat: matches 2..127 of consensus"
repeat_region    61016..61261
                 /note="MIR repeat: matches 12..255 of consensus"
repeat_region    61871..61914
                 /note="L2 repeat: matches 2666..2710 of consensus"
repeat_region    61921..61984
                 /note="32 copies 2 mer tt 71% conserved"
repeat_region    62213..6337
                 /note="L1PA5 repeat: matches 4980..6145 of consensus"
repeat_region    63376..64497
                 /note="L1P repeat: matches 3185..4306 of consensus"
repeat_region    64579..64700
                 /note="MER5A repeat: matches 7..131 of consensus"
repeat_region    66220..67231
                 /note="L1MB5 repeat: matches 5023..6138 of consensus"
repeat_region    67234..67320
                 /note="L1M4 repeat: matches 817..919 of consensus"
```

```
Query Match          3.0%;   Score 42.2;   DB 89;   Length 167721;
Best Local Similarity   47.8%;   Pred. No. 1.7;
Matches 122;   Conservative  0;   Mismatches 133;   Indels  0;   Gaps  0;

Qy     898 cagccatttctacattagtcagggccctatcctcttaacaaaagaccatatattcctcaat 957
            ||  ||| |||    | |    |||   ||  ||||||      |||   ||||| ||||||
Db  120185 CAAATTATTAAACAAAAAGTCATACTTTGAAACAAAACTTAATGATATAAAATCCTACC 120126

Qy     958 tcagcttaaggcaacagcccttacaagttctgtcagagagcatggagtcaaggct 1017
            | |   |||  |||  |     ||  | ||| | |   | || |  | |  ||||
Db  120125 AAAGTACAAGATCATTTCAGTTACTCAACTTTGTTGAAAAAAATGTAGAAAAAAT 120066

Qy    1018 gtgtttatgaataaacagactggtatgaagaatatccaactaccacctagctcttttgtt 1077
            ||  |||||| ||||||| ||| || |||||||| |   || |||| ||| | ||| |||
Db  120065 GTTTTATGTGTTAACAGCTCTTATGTACAGACAACAGAATCACTAATAATTATTTCTT 120006

Qy    1078 ttaaatggcttatgtattctttaattgggctgtatgacttaaaagaaactgcaggggaa 1137
            | |||| ||| ||| |||| ||||| || ||| |||| |||||||| ||| | ||||
Db  120005 TCACATATTTTTATATTTCTTAATAATTTCAAACTTTCTACAAATGATTATAATATTG 119946

Qy    1138 aaactcgggaagaa 1152
            ||| |   |||||| 
Db  119945 CAAATTTAAAAAGAA 119931
```

```
RESULT 15
LOCUS       CNS01DVU   174788 bp   DNA   HTG   08-DEC-2000
            CNS01DVU/c
DEFINITION  Homo sapiens chromosome 14 clone R-799P8, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL136039
VERSION     AL136039.3  GI:11611146
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174788)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Dec 9, 2000 this sequence version replaced gi:7708224.
            -------------- Genome Center
                Center: Genoscope / Centre National de Sequencage
                Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continue. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage, etc., even if efforts are made to eliminate these
            contaminating sequences. The following BAC sequence is oriented
            from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-299L17
            Downstream BAC (overlapping the SP6 end) : R-226P1 (AC=AL160192)
            -------------- Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 8.43x in Q20 bases; sum-of-contigs

            Overall quality chart :
            Range          bases
            0   -  9          136
            1   -  9            5
            10  - 19           54
            20  - 29          105
            30  - 39          552
            40  - 49         4248
            50  - 59         9727
            50  - 59         8979
            60  - 69        14516
            70  - 79        38520
            80  - 89        97946
            90  - 99

            Percentage of bases with a quality value >= 40 : 99 %.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..174788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-799P8"
                     /clone_lib="RPCI-11"
     STS             5811..5973
                     /note="matching EMBL:T95294
                     RHdb:RH66038
                     dbSTS:STS45968
                     Identified using the e-PCR software (G. Schuler)"
     STS             50407..50576
                     /note="matching EMBL:W84782
                     RHdb:RH66049
                     dbSTS:STS45979
                     Identified using the e-PCR software (G. Schuler)"
     STS             50558..50666
                     /note="matching EMBL:G41930
                     RHdb:RH101408
                     RHdb:RH94904
                     dbSTS:STS67895
                     Identified using the e-PCR software (G. Schuler)"
     STS             158308..158439
                     /note="matching EMBL:Z52777
                     RHdb:RH97871
                     RHdb:RH31321
                     dbSTS:STS28412
                     Identified using the e-PCR software (G. Schuler)"
BASE COUNT    53409 a   34948 c   35572 g  50858 t   1 others
ORIGIN
```

```
Query Match          3.0%;   Score 41.6;   DB 83;   Length 174788;
Best Local Similarity   57.8%;   Pred. No. 2.5;
Matches 74;   Conservative  0;   Mismatches 54;   Indels  0;   Gaps  0;
```

```
Qy    172 tatttctacccaatccagattgcacagtatgggttaagtcactacagcaagaatctaact 231
          |||||  ||         ||||||  |||||| ||        |||| ||    |  ||||
Db 172721 TATTTGTTAAGATTTCCAAATAGCACTTTACTCCTAAAAACACAAGAGAAATTATTCAGATT 172662

Qy    232 gaaaaacccctccatatagaggtatatgaaacagcaggagaacagggacaaaaacagcaag 291
          |||||| | |        ||  |||||| ||  |||| ||  ||||| |   || | ||
Db 172661 TAAAAATCACAGTTATGAACAAATCATTATCAAGCATCAAAATAGAGCCACAAACAGAAGGCAG 172602

Qy    292 cccaatga 299
          |  ||  |
Db 172601 CTGGATCA 172594
```

Search completed: May 25, 2001, 04:08:32
Job time: 6453 sec